

Thresholded Multiscale Gaussian Processes with Application to Bayesian Feature Selection for Massive Neuroimaging Data

Ran Shi^{*} and Jian Kang[†]

Abstract

Motivated by the needs of selecting important features for massive neuroimaging data, we propose a spatially varying coefficient model (SVCMS) with sparsity and piecewise smoothness imposed on the coefficient functions. A new class of nonparametric priors is developed based on thresholded multiscale Gaussian processes (TMGP). We show that the TMGP has a large support on a space of sparse and piecewise smooth functions, leading to posterior consistency in coefficient function estimation and feature selection. Also, we develop a method for prior specifications of thresholding parameters in TMGPs. Efficient posterior computation algorithms are developed by adopting a kernel convolution approach, where a modified square exponential kernel is chosen taking the advantage that the analytical form of the eigen decomposition is available. Based on simulation studies, we demonstrate that our methods can achieve better performance in estimating the spatially varying coefficient. Also, the proposed model has been applied to an analysis of resting state functional magnetic resonance imaging (Rs-fMRI) data from the Autism Brain Imaging Data Exchange (ABIDE) study, it provides biologically meaningful results.

Keywords: Thresholded multiscale Gaussian processes; Bayesian feature selection; Spatially varying coefficient models; Human brain mapping; Markov chain Monte Carlo

^{*}Ran Shi is Ph.D. Student, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA 30322., Jian Kang is Assistant Professor, Department of Biostatistics and Bioinformatics and Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA 30322.

[†]To whom correspondence should be addressed: jian.kang@emory.edu

1 Introduction

Recent advancements in biomedical imaging technologies have provided abundant information and extensive resources for researchers to learn the human brain and neurological diseases. A variety of imaging modalities, such as magnetic resonance imaging (MRI), diffusion tensor imaging (DTI), functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) have been developed to measure brain structures and functions from different perspectives, generating various large-scale spatially distributed measurements over a three dimensional (3D) space of the human brain. We refer to those massive spatial measurements of brain as neuroimages. This poses great opportunities and new challenges for neuroscientists and statisticians to develop efficient analytical methods that extract useful features from neuroimages to characterize the association between the brain activities and neurological diseases. To this end, regression analysis, a general and flexible modeling framework for studying the association among variables, has been investigated and considered as a powerful tool in the analysis of massive neuroimaging data, where neuroimages are modeled as outcome variables; and the disease status along with the clinical, biological and demographical information can all potentially be predictors.

A pioneer work using the regression model for the neuroimaging data is the mass univariate analysis (MUA). This approach fits a general linear model (GLM) at each spatial location in the brain (to which is referred as a voxel) and obtains massive test statistics over space to identify voxels/regions that are significantly associated with a specific covariate, which requires multiple comparisons correction. One standard procedure is to calculate the family-wise error rate (FWER) based on the random field theory for statistical parametric maps (Friston et al., 1995; Worsley et al., 2004; Lazar, 2008). Another approach is to control the false discovery rate (FDR) using the observed p-values (Benjamini and Yekutieli, 2001; Genovese et al., 2002). A major drawback of MUA is that the models do not borrow information from the spatial dependence across brain locations. In practice, the neuroimaging data are usually pre-processed by a spatial smoothing procedure using a kernel convolution approach. Performing MUA on these pre-smoothed data may lead to inaccuracy and low efficiency in terms of estimating and testing the covariate effects (Chumbley et al., 2009; Li et al., 2011). Recent development in adaptive smoothing methods for preprocessing (Yue

et al., 2010) and estimation (Polzehl and Spokoiny, 2000; Qiu, 2007; Tabelow et al., 2008b,a; Li et al., 2011; Wang et al., 2013) may improve the performance in terms of reducing noise and preserving features. It is especially powerful to detect delicate features such as jump discontinuities, which is one of the universal characteristics for neuroimaging data (Chan and Shen, 2005; Tabelow et al., 2008a,b; Chumbley et al., 2009).

To achieve a similar goal in the analysis of neuroimaging data, Zhu et al. (2014) recently developed a systematic modeling approach using a novel spatially varying coefficient model (SVCMM) which incorporates both spatial dependence and piecewise smooth covariate effects. General SVCMMs have been extensively investigated and developed for different applications in environmental health, epidemiology, ecology and geographical studies as demonstrated in Cressie and Cassie (1993); Diggle et al. (1998); Gelfand et al. (2003); Smith et al. (2002). The SVCMM encompasses a wide range of regression models with the outcome variable observed over space and the regression coefficients modeled as functions varying spatially. We refer to this type regression coefficients as spatially varying coefficient functions (SVCFFs). SVCFFs are commonly assumed to be smooth functions or ρ times continuously differentiable functions with $\rho \geq 1$ (we will not make this distinction throughout the rest of this paper unless noted). To model SVCFFs, smooth spatial processes are usually employed to characterize the dependence structure over space. A pure noise process is typically introduced to capture random variabilities, i.e. the “spatial nugget effect”. Zhu et al. (2014) extended the general SVCMMs by introducing jump discontinuities into the SVCFFs, making the model especially useful for neuroimaging data analysis. Based on stepwise multiscale estimating procedures and asymptotic Wald tests, Zhu et al. (2014)’s SVCMM also can identify the brain regions that are significantly associated with the given covariates, although it is not developed particularly for feature selection.

In this article, we aim to develop a Bayesian feature selection method for large-scale neuroimaging data that can directly select imaging features associated with covariates of interest. Regularization methods have been studied extensively for variable selection in regression models (Tibshirani, 1996; Fan and Li, 2001; Zou, 2006; Candes and Tao, 2007). Bayesian methods have also been developed based on various prior specifications. Mitchell and Beauchamp (1988) developed a prior model for linear model coefficients using the mix-

ture of a uniform distribution (slab) and a point mass at zero (spike), which is broadly referred to as the spike-and-slab type of priors. [George and McCulloch \(1993\)](#) proposed to use the scale mixture of two zero-mean Gaussian distributions and developed posterior computation algorithm based on Gibbs sampling. Relative works also include but not limited to [Ishwaran and Rao \(2005\)](#); [Liang et al. \(2008\)](#); [Park and Casella \(2008\)](#); [Hans \(2009\)](#); [Bondell and Reich \(2012\)](#); [Johnson and Rossell \(2012\)](#); [Armagan et al. \(2013\)](#); [Polson et al. \(2014\)](#); [Narisetty et al. \(2014\)](#). Most of these priors were initially introduced for independent regression coefficients. In light of the needs of integrating complex data structure in many applications, recent development of Bayesian variable selection incorporates dependence structures into the prior model. [Li and Zhang \(2010\)](#) assumed that covariates lay on an undirected graph and used the Ising prior to incorporate this information to the model space and applied this method to analyze the genomics data. For the modeling of spatial data, Markov random field (MRF) is one of the commonly used priors for regression coefficients. For instance, [Smith et al. \(2003\)](#); [Smith and Fahrmeir \(2007\)](#) applied this type of priors to fMRI data analyses. For the analysis of physical activity and environmental health data, [Reich et al. \(2010\)](#) developed an multivariate SVCMM along with a Bayesian variable selection procedure to identify important SVCMMs, using the spike-and-slab prior. Their focus, however, was on distinguishing covariate effects that were zero constant, nonzero constant and spatially varying instead of selecting features within the varying coefficient functions.

As the aforementioned variable selection methods cannot be directly applied to identify important features for massive neuroimaging data in the SVCMM framework. To fill this gap, we develop a novel Bayesian nonparametric prior model for the SVCMM. We refer to it as the thresholded multiscale Gaussian process (TMGP). The TMGP prior is constructed by thresholding a multiscale Gaussian process, which is a combination of two types of GPs: a global GP to account for the entire domain spatial dependence and a local GP to accommodate the regional fluctuations. Thus, the TMGP can characterize important common features of the neuroimaging data, including sparsity, global spatial dependence, piecewise smoothness, edge effects and jumps. The proposed TMGP prior enjoys the large support property, leading to posterior strong consistency in estimation and feature selection for the sparse and piecewise smooth SVCMMs. We also develop efficient MCMC posterior compu-

tation algorithms based on a kernel convolution approach. A special choice of the kernel function enables the computation scalable to an ultra-high dimensional case.

The remaining parts of the article is organized as follows: In Section 2, we first introduce the SVCs for neuroimaging data analysis and particularly discuss conditions on SVCs in the proposed model. Then we present the construction of TMGP which serves as a prior model for SVCs. In Section 3, we study the theoretical properties of TMGP and the proposed SVCs. In Section 4, we develop an efficient and scalable posterior computation algorithm based on a kernel convolution approach. We evaluate the performance of proposed method via simulation studies and analyze the ABIDE data in Section 5. We conclude our work with a brief discussion on the future work in Section 6.

2 Feature selection within the spatially varying coefficient functions

We start with general notations and definitions. Denote by \mathbb{R}^p a p -dimensional real Euclidean space for any $p \geq 1$. For any $\boldsymbol{\beta} \in \mathbb{R}^p$, write $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$, define $\|\boldsymbol{\beta}\|_\infty = \max_{1 \leq k \leq p} |\beta_k|$, $\|\boldsymbol{\beta}\|_1 = \sum_{k=1}^p |\beta_k|$ and $\|\boldsymbol{\beta}\|_2 = \sqrt{\sum_{k=1}^p \beta_k^2}$. Denote by $\mathcal{R} \subset \mathbb{R}^3$ a compact region in the standard brain space. Let $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{R}$ be a set of spatial locations where brain signals are measured. An empirical measure on \mathcal{R} induced by $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ is defined as $\mathbb{P}_n(d\mathbf{s}) = \frac{1}{n} \sum_{i=1}^n I[\mathbf{s}_i \in d\mathbf{s}]$, where the indicator function $I[\mathcal{A}] = 1$ if event \mathcal{A} occurs, $I[\mathcal{A}] = 0$, otherwise. For a scalar-valued function $\beta(\cdot) : \mathcal{R} \mapsto \mathbb{R}$, define $\|\beta(\cdot)\|_\infty = \sup_{\mathbf{s} \in \mathcal{R}} |\beta(\mathbf{s})|$ and $\|\beta(\cdot)\|_1 = \int_{\mathbf{s} \in \mathcal{R}} |\beta(\mathbf{s})| \mathbb{P}_n(d\mathbf{s})$. For a p -dimensional vector-valued function $\boldsymbol{\beta}(\mathbf{s}) = [\beta_1(\mathbf{s}), \dots, \beta_p(\mathbf{s})]^T : \mathcal{R} \mapsto \mathbb{R}^p$, define $\|\boldsymbol{\beta}(\cdot)\|_{1,\infty} = \max_{1 \leq k \leq p} \|\beta_k(\cdot)\|_1$. Denote by $\mathcal{C}(\mathcal{R})$ a collection of all the continuous functions defined on \mathcal{R} . Let $D^\alpha \beta$ be a partial derivative operator on function $\beta(\cdot)$ (given its existence) which is given by $\frac{\partial^{\|\alpha\|_1} \beta}{\partial s_1^{\alpha_1} \dots \partial s_d^{\alpha_d}}$ for $\alpha \in \mathbb{R}^p$. Denote by $\mathcal{C}^\rho(\mathcal{R})$ a set of functions β defined on \mathcal{R} with continuous partial derivatives $D^\alpha \beta$ for all α such that $\|\alpha\|_1 \leq \rho$.

2.1 The spatially varying coefficient model for neuroimaging data

Suppose the data set consists of m subjects. For each subject j , let $y_j(\mathbf{s})$ be the brain signal measured from a certain imaging modality at location $\mathbf{s} \in \mathcal{R}$; and there are also p covariates are collected, denoted $\mathbf{x}_j = (x_{j1}, \dots, x_{jp})^T$, for $j = 1, \dots, m$. The spatially varying coefficient

model (SVCMM) for neuroimaging data is given by

$$y_j(\mathbf{s}) = \mathbf{x}_j^T \boldsymbol{\beta}(\mathbf{s}) + e_j(\mathbf{s}), \quad (1)$$

where $\boldsymbol{\beta}(\mathbf{s}) = [\beta_1(\mathbf{s}), \dots, \beta_p(\mathbf{s})]^T$ is the spatially varying coefficient function (SVCF) defined on \mathcal{R} . It characterizes associations between covariates and imaging outcomes. To be more specific, $\beta_k(\mathbf{s})$ ($k = 1, \dots, p$) quantifies the k th covariate effects at brain location \mathbf{s} . The independent error process $e_j(\mathbf{s})$ is assumed to be spatially homogeneous across the whole brain for each subject. In our model, we simply assume that $e_j(\mathbf{s}) \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ for all subjects at all locations.

For neuroimaging data from commonly used imaging modalities, only outcomes at locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ are observed. At these locations, the SVCMM proposed in (1) for the recorded neuroimaging data can be expressed as

$$[\mathbf{y}(\mathbf{s}_i) \mid \boldsymbol{\beta}(\mathbf{s}_i), \sigma^2] \sim N(\mathbf{X}\boldsymbol{\beta}(\mathbf{s}_i), \sigma^2 \mathbf{I}_m) \quad (2)$$

independently for all $i = 1, \dots, n$, where $\mathbf{y}(\mathbf{s}_i) = [y_1(\mathbf{s}_i), \dots, y_m(\mathbf{s}_i)]^T$, $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_m]^T$ and $\mathbf{e}(\mathbf{s}_i) = [e_1(\mathbf{s}_i), \dots, e_m(\mathbf{s}_i)]^T$. For simplicity, denote by $\mathcal{Y} = \{\mathbf{y}(\mathbf{s}_i)\}_{i=1}^n$ an $m \times n$ matrix recoding all the neuroimaging outcomes involved in the study.

In neuroimaging studies, there exists a natural region partition of the whole brain domain \mathcal{R} into bounded connected sets $\mathcal{R}_1, \dots, \mathcal{R}_G$ with non-empty interiors, such that $\mathcal{R} = \cup_{g=1}^G \mathcal{R}_g$, $\mathcal{R}_g \cap \mathcal{R}_{g'} = \emptyset$, $\forall g \neq g'$. In many cases, one can utilize \mathcal{R}_g as neuroanatomical regions from commonly used labeling systems such as the Automated Anatomical Labeling (AAL) (Tzourio-Mazoyer et al., 2002). For region of interest (ROI) based analysis, each ROI is one parcellated region. For seed-based region-level analysis, \mathcal{R}_g can be regarded as the clusters showing strong functional connectivities based on preliminary results. In some voxelwise analysis with no regional information to be incorporated, we can simply consider each voxel (a 3D cubic) as a region and the centers of voxels as observed brain locations $\mathbf{s}_1, \dots, \mathbf{s}_n$.

To utilize the proposed SVCMM for analysis of neuroimaging data and feature selection, we state our assumptions for the SVCF in model (1). Specifically, we work with piecewise smooth (ρ -times continuously differentiable to be accurate) functions with structured sparsity, which can be mathematically expressed as follows: for any function $\beta(\mathbf{s})$ defined on \mathcal{R} to be a

varying coefficient function within model (1), $\beta(\mathbf{s})$ must satisfy that:

(C1) there exists an index set $I_1 \subset \{1, \dots, G\}$, such that $\beta(\mathbf{s}) \times I[\mathbf{s} \in \overline{\mathcal{R}}_g] \in \mathcal{C}^\rho(\overline{\mathcal{R}}_g)$ where

$\overline{\mathcal{R}}_g$ is the closure of R_g , for all $g \in I_1$ with $\rho = \lfloor \frac{d}{2} \rfloor + 1$;

(C2) for any $g \in I_1$, $\beta(\mathbf{s})$ is bounded away from zero, that is,

$$\lambda = \inf_{\mathbf{s} \in \cup_{g \in I_1} \mathcal{R}_g} |\beta(\mathbf{s})| > 0;$$

(C3) let $I_0 = \{1, \dots, G\} \setminus I_1$, then $\beta(\mathbf{s}) = 0$ for all $\mathbf{s} \in \cup_{g \in I_0} \mathcal{R}_g$.

The conditions on the SVCFs can be interpreted as follows: (C1) demonstrates that the functions are smooth within each brain region, which implies more homogeneous covariate effects; (C2) indicates the jump discontinuities at the boundaries of brain regions; (C3) introduces sparsity into each SVCF in model (1) and restricts the sparsity structure at the regional level.

We introduce the notation \mathcal{P} to represent a set of functions satisfying (C1)–(C3) defined on \mathcal{R} . The definition \mathcal{P} carries the information of R_1, \dots, R_G as well as λ and I_1 . For any function $\beta(\mathbf{s}) \in \mathcal{P}$, denote by $I_1\{\beta(\mathbf{s})\}$ the index of nonzero sets as given in (C1) and by $\Lambda\{\beta(\mathbf{s})\}$ the value λ defined in (C2). In the same vein, define a set of p -dimensional vector-valued functions $\mathbf{P} = \{\beta(\mathbf{s}) = [\beta_1(\mathbf{s}), \dots, \beta_p(\mathbf{s})]^\top : \beta_k(\mathbf{s}) \in \mathcal{P}, k = 1, \dots, p\}$.

2.2 The thresholded multiscale Gaussian process priors

2.2.1 Construction of the prior

A Gaussian process (GP) can be regarded as a probabilistic measure on certain functional spaces, making it as popular prior models in Bayesian nonparametric data analysis. In general, the GP prior, denoted by $\mathcal{GP}[\mu(\cdot), C(\cdot, \cdot)]$, is determined by its mean function $\mu : \mathcal{R} \mapsto \mathbb{R}$ and the covariance kernel function $C : \mathcal{R} \times \mathcal{R} \mapsto \mathbb{R}$. A draw $\beta(\cdot) \sim \mathcal{GP}[\mu(\cdot), C(\cdot, \cdot)]$ is a function defined on \mathcal{R} such that any finite collection of its function values are jointly multivariate Gaussian. To be specific, for any choices of $\mathbf{s}_1, \dots, \mathbf{s}_n \in \mathcal{R}$, $[\beta(\mathbf{s}_1), \dots, \beta(\mathbf{s}_n)]^\top \sim N(\boldsymbol{\mu}, \mathbf{C})$ with $\boldsymbol{\mu} = [\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_n)]^\top$ and $\mathbf{C} = \{C(\mathbf{s}_i, \mathbf{s}_j)\}_{1 \leq i \leq n, 1 \leq j \leq n}$. The boundedness and smoothness of random functions generated from GP are determined through the covariance kernel function. Typical choices for the covariance kernel functions include but not limited to the rational quadratic kernel, Matérn class of kernels, the square exponential kernel. [Stein](#)

(1999); Williams and Rasmussen (2006) contain more detailed discussions on the covariance functions.

To enable detailed feature selections within the SVCFs $\beta_k(\mathbf{s}) \in \mathcal{P}$ ($k = 1, \dots, p$) in model (1), we develop the thresholded multiscale Gaussian process (TMGP) prior, which can be represented as follows: $\beta(\mathbf{s}) \sim \mathcal{TMGP}[\tau^2, \theta^2, \lambda, \kappa(\cdot, \cdot)]$ implies that

$$\beta(\mathbf{s}) = \tilde{\beta}(\mathbf{s})I_\lambda[\tilde{\beta}(\mathbf{s})], \quad (3)$$

$$\tilde{\beta}(\mathbf{s}) = \gamma(\mathbf{s}) + \epsilon(\mathbf{s}), \quad (4)$$

$$\gamma(\mathbf{s}) \sim \mathcal{GP}[0, \tau^2 \kappa(\mathbf{s}, \mathbf{s}')], \quad (5)$$

$$\epsilon(\mathbf{s}) \sim \mathcal{GP}[0, \theta^2 \chi(\mathbf{s}, \mathbf{s}')] \quad (6)$$

for all $\mathbf{s} \in \mathcal{R}$, where

$$I_\lambda[\tilde{\beta}(\mathbf{s})] = \sum_{g=1}^G I \left[\mathbf{s} \in \mathcal{R}_g : \inf_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}(\mathbf{s})| > \lambda \right] \quad (7)$$

is the thresholding function that generate region-wise sparse features; $\lambda > 0$ is the thresholding parameter; $\tau^2 > 0$ and $\theta^2 > 0$ are variance parameters in the GPs; $\kappa(\cdot, \cdot) : \mathcal{R} \times \mathcal{R} \mapsto \mathbb{R}$ is a kernel correlation function and $\chi(\mathbf{s}, \mathbf{s}')$ is constructed from κ as $\chi(\mathbf{s}, \mathbf{s}') = \sum_{g=1}^G \kappa(\mathbf{s}, \mathbf{s}') \times I[\mathbf{s}, \mathbf{s}' \in \mathcal{R}_g]$.

The thresholding construction of $\beta(\mathbf{s})$ from a particular multiscale Gaussian process (Dunson and Fox, 2012, mGP) $\tilde{\beta}(\mathbf{s})$ introduces sparsity and enables feature selection. For voxel-wise analysis without regional information, the thresholding function in (7) can be simplified as $I_\lambda[\tilde{\beta}(\mathbf{s})] = I[|\tilde{\beta}(\mathbf{s})| > \lambda]$, where \mathbf{s} denotes the center of a voxel. The mGP $\tilde{\beta}(\mathbf{s})$ in our prior is a combination of one “global” GP, $\gamma(\mathbf{s})$, which captures the general dependence structures across the whole brain domain and one “local” GP, $\epsilon(\mathbf{s})$, reflecting the dependence and variabilities within each parcellated brain region. This multiscale construction can also naturally generate jumping discontinuities on the boundaries of the brain regions.

We illustrate the procedure to sample an SVCf from our TMGP priors on a two-dimensional square region in Figure 1, where the dashed lines partition the whole region into four equally spaced sub-regions. $\gamma(\mathbf{s})$ is smooth over the whole region; $\epsilon(\mathbf{s})$ is smooth within each sub-region but has distinct jumps on the boundaries. As we may see from Figure 1, one particular issue is that the generating processes, $\gamma(\mathbf{s})$ and $\epsilon(\mathbf{s})$, could vary substan-

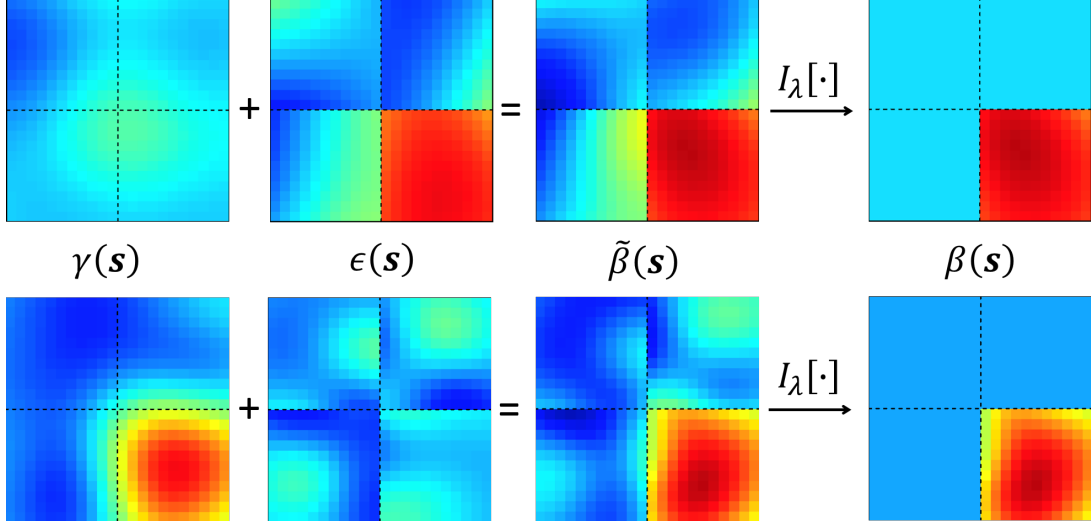


Figure 1: Sample SVCFs from TMGP prior. Top row: weak global signals with strong local signals; Bottom row: strong global signals with weak local signals.

tially while results in similar sparse SVCFs. This phenomenon indicates weak identifiability in global and local GPs in the TMGP prior model. To resolve this issue, we impose the certain constraint on the marginal variance of the local GP $\epsilon(\mathbf{s})$. i.e., θ^2 . This implies that we assume that the sampled SVCF is constructed from strong global signals and weak local signals.

3 Theoretical results

We first introduce two sets of extra conditions in addition to the conditions (C1)-(C3) for the SVCFs. The design matrix \mathbf{X} as defined in (2) satisfies:

(X1) Let d_{\min} and d_{\max} be the smallest and largest eigenvalues of $\frac{1}{m}\mathbf{X}^\top\mathbf{X}$, then $0 < d_{\min} < d_{\max} < \infty$.

For the kernel correlation functions $\kappa(\cdot, \cdot)$ in our proposed TMGP priors, we introduce the following condition:

(K1) $\kappa(\mathbf{s}, \mathbf{s}') = \prod_{j=1}^d K_j(\|\mathbf{s} - \mathbf{s}'\|)$ for some nowhere zero, continuous, symmetric density function (up to a normalization constant) K_j defined on \mathbb{R} .

(K2) $\kappa(\mathbf{s}, \cdot)$ has continuous partial derivatives up to order $2\rho + 2$ where $\rho = \lceil \frac{d}{2} \rceil + 1$.

Theorem 1. Consider an arbitrary SVCF $\beta^0(\mathbf{s}) \in \mathcal{P}$, i.e. $\beta^0(\mathbf{s})$ satisfies conditions (C1)-(C3). Let $\lambda_0 = \Lambda[\beta^0(\mathbf{s})]$ and suppose the partition number $G < \infty$. If $0 < \lambda < \lambda_0, 0 < \tau^2, \theta^2 < \infty$ and the kernel function κ satisfies (K1), then the proposed prior

$$\beta(\mathbf{s}) \sim \mathcal{TMGP}[\tau^2, \theta^2, \lambda, \kappa(\cdot, \cdot)]$$

given in (3)-(6) satisfies that

$$\Pi(\|\beta(\mathbf{s}) - \beta^0(\mathbf{s})\|_\infty < \varepsilon) > 0 \quad \text{for all } \varepsilon > 0$$

Theorem 1 demonstrates that the proposed TMGP prior assign positive measures to arbitrarily small neighborhoods of all elements within \mathcal{P} , the family of spatially varying coefficient functions defined in our model (1). This property is essential, especially for Bayesian nonparametric priors, since it is necessary for appropriate posterior behaviors and can not be guaranteed in many cases.

Theorem 2. Suppose that our observed data satisfies $\mathbf{y}(\mathbf{s}_i) \sim N(\mathbf{X}\beta^0(\mathbf{s}_i), \sigma^2 \mathbf{I}_m)$ independently following the notations in (2), with a known σ^2 and a fixed p -dimensional SVCF $\beta^0(\mathbf{s}) = [\beta_1^0(\mathbf{s}), \dots, \beta_p^0(\mathbf{s})]^T$, $p < m$, defined on \mathcal{R} . Suppose $\beta_k^0(\mathbf{s})$ satisfies conditions (C1)-(C3), i.e. $\beta^0(\mathbf{s}) \in \mathbf{P}$, with $G < \infty$; \mathbf{X} satisfies condition (X1); the kernel function in the TMGP priors satisfies (K1) and (K2). For all $\varepsilon > 0$, if $\frac{\sigma^2}{m} < \frac{\varepsilon^2 d_{\min}}{8 \log 2}$ and each dimension of $\beta(\mathbf{s})$ follows a TMGP prior independently satisfying the conditions in Theorem 1, then the posterior distribution satisfies that

$$\Pi[U_\varepsilon^c \mid \mathbf{y}(\mathbf{s}_1), \dots, \mathbf{y}(\mathbf{s}_n)] \rightarrow 0,$$

as $n \rightarrow \infty$ in $P_{\beta^0}^n$ probability, where $U_\varepsilon = \{\beta(\mathbf{s}) \in \mathbf{P} : \|\beta(\mathbf{s}) - \beta^0(\mathbf{s})\|_{1,\infty} < \varepsilon\}$.

Theorem 2 justifies the posterior consistency of the proposed TMGP prior given model (1) under the infill asymptotic framework. It implies that, if a ground truth of the SVCFs exists and the data is generated accordingly, then the posterior distribution of $\beta(\mathbf{s})$ can be concentrated to an arbitrarily small $\|\cdot\|_{1,\infty}$ neighborhood around the truth as the number of spatial locations goes to infinity. The conditions of this theory also imply that a small ratio between the number of subjects and the variance of pure noise, i.e., σ^2/m , is also important to guarantee a good performance of our method. One limitation of Theorem 2 is that it does not apply to the voxel level analysis where $G = n \rightarrow \infty$. However, this type of analysis generally works well empirically.

Although the $\|\cdot\|_{1,\infty}$ norm is not common in Bayesian asymptotic literatures, a direct result based on Theorem 2 is the element-wise posterior consistency for $\boldsymbol{\beta}(\mathbf{s})$ under the commonly used empirical $\|\cdot\|_1$ norm for a fixed design of spatial locations \mathbf{s}_i (Ghosal et al., 2006).

Corollary 1. *Under the same assumptions and conditions in Theorem 2, for all $\varepsilon > 0$, if $\sigma^2/m < \frac{\varepsilon^2 d_{\min}}{8 \log 2}$, then the posterior distribution satisfies that for all $k = 1, \dots, p$,*

$$\Pi [U_{\varepsilon,k}^c \mid \mathbf{y}(\mathbf{s}_1), \dots, \mathbf{y}(\mathbf{s}_n)] \rightarrow 0$$

as $n \rightarrow \infty$ in $P_{\boldsymbol{\beta}_0}^n$ probability, where $U_{\varepsilon,k} = \{\boldsymbol{\beta}(\mathbf{s}) \in \mathcal{P} : \|\boldsymbol{\beta}(\mathbf{s}) - \boldsymbol{\beta}_k^0(\mathbf{s})\|_1 < \varepsilon\}$.

Of note, in neuroimaging studies, \mathbf{s}_i are usually fixed 3D grid points, thus we do not consider the $\|\cdot\|_1$ norms with regard to random measures for \mathbf{s} when proving posterior consistency. For a fixed design within a finite domain \mathcal{R} (the volume of brain is limited), another useful direction is to show posterior consistency under the $\|\cdot\|_1$ norm with regard to the Lebesgue measure. We see this as a potential extension to the theory development in the future work.

4 Posterior Inferences

4.1 Model Representation

Now consider the SVCMM defined in (1). For the p -dimensional multivariate spatially varying coefficient function, $\boldsymbol{\beta}(\mathbf{s}) = [\beta_1(\mathbf{s}), \dots, \beta_k(\mathbf{s})]^T$, we assume that

$$\beta_k(\mathbf{s}) \sim \mathcal{TMGCP}[\tau_k^2, \theta^2, \lambda_k, \kappa(\cdot, \cdot)],$$

with $\kappa(\cdot, \cdot)$ being a smooth kernel function. This specification implies that the global processes (5) have distinct flexible variance parameters τ_k^2 , while the local fluctuation processes (6) have a small fixed marginal variance θ^2 .

Based on the prior specification for $\boldsymbol{\beta}(\mathbf{s})$, we have that $\beta_k(\mathbf{s}) = \tilde{\beta}_k(\mathbf{s})I_{\lambda_k}[\tilde{\beta}_k(\mathbf{s})]$ for $k = 1, \dots, p$, where

$$\tilde{\beta}_k(\mathbf{s}) = \gamma_k(\mathbf{s}) + \epsilon_k(\mathbf{s}), \tag{8}$$

where $\gamma_k(\mathbf{s}) \sim \mathcal{GP}[0, \tau_k^2 \kappa(\mathbf{s}, \mathbf{s}')]]$ and $\epsilon_k(\mathbf{s}) \sim \mathcal{GP}\left(0, \sum_{g=1}^G \theta^2 \kappa(\mathbf{s}, \mathbf{s}') \times I[\mathbf{s}, \mathbf{s}' \in R_g]\right)$. For global GPs: $\gamma_k(\mathbf{s})$ in (8), its Karhunen-Loève (KL) expansion can be expressed as

$$\gamma_k(\mathbf{s}) = \sum_{l=1}^{\infty} \varphi_l(\mathbf{s}) u_{kl}, \tag{9}$$

where $u_{kl} \sim N(0, \tau_k^2 \zeta_l)$ independently with $\zeta_l > 0$ such that $\sum_{l=1}^{\infty} \zeta_l \varphi_l(\mathbf{s}) \varphi_l(\mathbf{s}') = \kappa(\mathbf{s}, \mathbf{s}')$ and that $\int \varphi_l(\mathbf{s}) \varphi_{l'}(\mathbf{s}) d\mathbf{s} = 0, \forall l \neq l'$ based on the Mercer's Theorem (Williams and Rasmussen, 2006). In practice, we truncate the infinite sum in (9) into L terms such that $\sum_{l=1}^L \zeta_l / \sum_{l=1}^{\infty} \zeta_l$ is close to 1.

This implies a model representation of the proposed SVCM along with the TMGP prior specifications. To be more specific, the neuroimaging signal $y_j(\mathbf{s})$ on locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ can be modeled through latent mGPs: $\tilde{\boldsymbol{\beta}}(\mathbf{s}) = [\tilde{\beta}_1(\mathbf{s}), \dots, \tilde{\beta}_p(\mathbf{s})]^T$ and the truncated KL expansion coefficients $\{\mathbf{u}_k\}_{k=1}^p$ with $\mathbf{u}_k = [u_{k1}, \dots, u_{kL}]^T$ by integrating out the local GPs $\epsilon_k(\mathbf{s})$ in (8), which is given by

$$[y_j(\mathbf{s}_i) \mid \tilde{\boldsymbol{\beta}}(\mathbf{s}_i), \sigma^2] \sim N\left(\mathbf{x}_j^T g_{\boldsymbol{\lambda}}[\tilde{\boldsymbol{\beta}}(\mathbf{s}_i)], \sigma^2\right), \quad (10)$$

$$[\{\tilde{\beta}_k(\mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{R}_g} \mid \mathbf{u}_k] \sim N(\boldsymbol{\varphi}_g \mathbf{u}_k, \theta^2 K_g), \quad (11)$$

$$u_{kl} \sim N(0, \zeta_l \tau_k^2), \quad (12)$$

for $j = 1, \dots, m$, $i = 1, \dots, n$, $k = 1, \dots, p$, $g = 1, \dots, G$ and $l = 1, \dots, L$, where $N(\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2 , $\boldsymbol{\varphi}_g = \{\boldsymbol{\varphi}(\mathbf{s}_i)^T\}_{\mathbf{s}_i \in \mathcal{R}_g}$ with $\boldsymbol{\varphi}(\mathbf{s}_i) = [\varphi_1(\mathbf{s}_i), \dots, \varphi_L(\mathbf{s}_i)]^T$ and $K_g = \{\kappa(\mathbf{s}_i, \mathbf{s}_{i'})\}_{\mathbf{s}_i, \mathbf{s}_{i'} \in \mathcal{R}_g}$ is a correlation matrix. The p -dimensional vector value functional operator $g_{\boldsymbol{\lambda}}(\cdot)$ is defined on the domain of all functions in \mathcal{P} , which is given by

$$g_{\boldsymbol{\lambda}}[\tilde{\boldsymbol{\beta}}(\mathbf{s})] = \left[\tilde{\beta}_1(\mathbf{s}) I_{\lambda_1}[\tilde{\beta}_1(\mathbf{s})], \dots, \tilde{\beta}_p(\mathbf{s}) I_{\lambda_p}[\tilde{\beta}_p(\mathbf{s})] \right]^T,$$

with $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_p]^T$.

4.2 Hyper Prior Specifications

We assign conjugate priors to variance parameters σ^2 and τ_k^2 in the TMGP model, i.e. $\sigma^2 \sim \text{Inv-Ga}(v, w)$ and $\tau_k^2 \stackrel{\text{iid}}{\sim} \text{Inv-Ga}(v, w)$, where $\text{Inv-Ga}(v, w)$ represents an inverse gamma distribution with shape v and rate w . We fix $\theta^2 = 1$ to restrict local deviations in order to sort out the weakly identifiability in the model, especially for the case that a small number of observations are recorded in each region.

We develop a data-driven method to specify the prior of thresholding parameters $\lambda_k, k = 1, \dots, p$. We consider the log full conditional of λ_k which is given by

$$\log \pi[\lambda_k \mid \mathcal{Y}, \tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}, \sigma^2] = \ell[\lambda_k; \mathcal{Y}, \tilde{\boldsymbol{\beta}}_k, \boldsymbol{\beta}_{-k}] / \sigma^2 + C,$$

where C is a constant, $\tilde{\beta}_k = [\tilde{\beta}_k(\mathbf{s}_1), \dots, \tilde{\beta}_k(\mathbf{s}_n)]^T$, $\beta_k = [\beta_k(\mathbf{s}_1), \dots, \beta_k(\mathbf{s}_n)]^T$, $\beta_{-k} = [\beta_1, \dots, \beta_{k-1}, \beta_{k+1}, \dots, \beta_p]^T$, and

$$\ell(\lambda_k) := \ell[\lambda_k; \mathcal{Y}, \tilde{\beta}_k, \beta_{-k}] = \sum_{i=1}^n \omega_k(\mathbf{s}_i) I_{\lambda_k}[\tilde{\beta}_k(\mathbf{s}_i)], \quad (13)$$

with $\omega_k(\mathbf{s}) = \sum_{j=1}^m \tilde{\beta}_k(\mathbf{s}) x_{jk} [2y_{j,-k}(\mathbf{s}) - \tilde{\beta}_k(\mathbf{s}) x_{jk}]$ and $y_{j,-k}(\mathbf{s}) = y_j(\mathbf{s}) - \sum_{j' \neq k} x_{jj'} \beta_{j'}(\mathbf{s})$. The function $\ell(\lambda_k)$ is flat when λ_k is around zero and dramatically decreases when λ_k is greater a certain value, to which we refer as a “turning point”. It should be close to the true threshold. Figure 2 shows the profiles of $\ell(\lambda_k)$ for a model with three SVCFs on a space of 900 locations from 50 simulated datasets. The true thresholds $\lambda_k = k + 1$ for $k = 1, 2, 3$. The turning points in the profiles of $\ell(\lambda_k)$ are all around the true thresholds. Thus, we can specify the priors of λ_k according to $\ell(\lambda_k)$. In practice, we need to provide rough estimates of $\tilde{\beta}_k$ and β_{-k} in order to evaluate $\ell(\lambda_k)$ before posterior inferences. We consider an SVMCM with smoothed SVCFs approximated by the truncated K-L expansion, where we compute the ordinary least squares (OLS) of the coefficients, i.e.

$$\{\hat{w}_{lk}\}_{l=1}^p = \arg \min_{\{w_{lk}\}} \sum_{j=1}^m \sum_{i=1}^n \left(y_j(\mathbf{s}_i) - \sum_{k=1}^p \sum_{l=1}^L x_{jk} \varphi_l(\mathbf{s}_i) w_{lk} \right)^2.$$

Then both $\tilde{\beta}_k(\mathbf{s})$ and $\beta_k(\mathbf{s})$ can be approximated by $\hat{\beta}_k(\mathbf{s}) = \sum_{l=1}^L \varphi_l(\mathbf{s}) \hat{w}_{lk}$. Thus, we replace β_k and $\tilde{\beta}_k$ by $\hat{\beta}_k = [\hat{\beta}_k(\mathbf{s}_1), \dots, \hat{\beta}_k(\mathbf{s}_n)]^T$ and $\hat{\beta}_{-k} = [\hat{\beta}_1, \dots, \hat{\beta}_{k-1}, \hat{\beta}_{k+1}, \dots, \hat{\beta}_p]^T$, respectively, in (13). Write $\hat{\ell}(\lambda_k) = \ell(\lambda_k; \mathcal{Y}, \hat{\beta}_k, \hat{\beta}_{-k})$.

We propose to assign uniform priors to λ_k , i.e. $\lambda_k \sim \text{Unif}(c_k - h_k, c_k + h_k)$, where the half range h_k and center c_k can be determined based on the profile of $\hat{\ell}(\lambda_k)$. More specifically, we evaluate $\hat{\ell}(\lambda_k)$ on a set of grid points $\{\lambda_k^{(1)}, \dots, \lambda_k^{(G)}\}$, denoted $\{\ell_k^{(1)}, \dots, \ell_k^{(G)}\}$. Given an interval (a, b) , define the sample correlation between λ_k and $\hat{\ell}(\lambda_k)$ within (a, b) as

$$\hat{\rho}(a, b) = \frac{\sum_{\lambda_k^{(g)} \in (a, b)} (\lambda_k^{(g)} - \bar{\lambda}_k) (\ell_k^{(g)} - \bar{\ell}_k)}{\sqrt{\sum_{\lambda_k^{(g)} \in (a, b)} (\lambda_k^{(g)} - \bar{\lambda}_k)^2} \sqrt{\sum_{\lambda_k^{(g)} \in (a, b)} (\ell_k^{(g)} - \bar{\ell}_k)^2}}$$

with $\bar{\lambda}_k = \sum_{\lambda_k^{(g)} \in (a, b)} \lambda_k^{(g)} / M(a, b)$, $\bar{\ell}_k = \sum_{\lambda_k^{(g)} \in (a, b)} \ell_k^{(g)} / M(a, b)$ and $M(a, b) = \sum_{g=1}^G I[\lambda_k^{(g)} \in (a, b)]$. And define

$$\tilde{c}_k(h) = \min\{\lambda_k^{(g)} : |\hat{\rho}(\lambda_k^{(g)} - h, \lambda_k^{(g)} + h)| > \zeta_k\},$$

where ζ_k is determined by the rejection region of Pearson correlation test. Given $h > 0$, $\tilde{c}_k(h)$

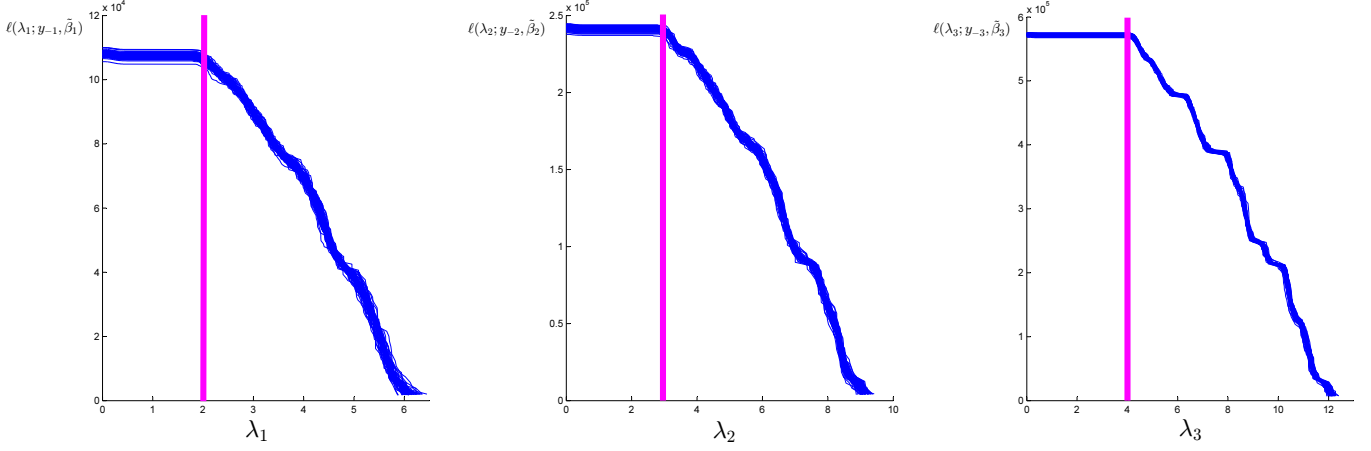


Figure 2: Simulated $\widehat{\ell}(\lambda_k)$ from 50 synthetic datasets: ground truth ($\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 4$) are marked in the figures.

represents the location that $\widehat{\ell}(\lambda_k)$ and λ_k have no significant correlation. Then we specify

$$h_k = \min\{h : |\widehat{\rho}(\widetilde{c}_k(h) - h, \widetilde{c}_k(h) + h)| > \zeta_k\}, \quad \text{and} \quad c_k = \widetilde{c}_k(h_k).$$

This leads to an informative prior range $[h_k - c_k, h_k + c_k]$ for λ_k with a high probability to cover the turning point of $\ell(\lambda_k)$.

4.3 Kernel Expansion for Massive Data Analysis

Theoretically, the KL expansion for the GP with kernel function $\kappa(\cdot, \cdot)$ relies on solving the integral equation $\int \kappa(\mathbf{s}, \mathbf{s}') \varphi(\mathbf{s}) d\mathbf{s} = \zeta_l \varphi(\mathbf{s}')$, which might not admit analytical solutions. Empirically, the expansion is often achieved by calculating the eigenvalues and eigenvectors of the $n \times n$ correlation matrix $K_n = \{\kappa(\mathbf{s}_i, \mathbf{s}_{i'})\}_{1 \leq i, i' \leq n}$ on a set of pre-specified locations. However, in the analysis of massive neuroimaging data that can involve a very large number (n can be hundreds of thousands) of brain locations, it is computationally infeasible to perform eigen decompositions on K_n . To solve this issue, we introduce the modified square exponential kernel

$$\kappa(\mathbf{s}, \mathbf{s}') = \exp\{-a\|\mathbf{s}\|_2^2 - a\|\mathbf{s}'\|_2^2 - b\|\mathbf{s} - \mathbf{s}'\|_2^2\}, \quad a, b > 0 \quad (14)$$

with a relatively small value for a as a numerical approximation to the square exponential kernel when dealing with massive neuroimaging data. The major benefit of this kernel function is that it has analytically tractable expansion. The detailed properties of this

kernel is summarized in Proposition 1.

Proposition 1. *For a specific $l \in \{1, \dots, \infty\}$, define series $\{k_i\}_{i=0}^d$, $\{l_i\}_{i=0}^d$ and $\{m_i\}_{i=1}^d$ as follows*

$$\begin{aligned} k_i &= \left\{ k_i \in \mathbb{N}^0 : \binom{k_i + d - i - 1}{d - i} \leq l_i \leq \binom{k_i + d - i}{d - i} - 1 \right\}, \quad 0 \leq i \leq d - 1, \quad k_d = 0, \\ l_0 &= l - 1, \quad l_i = l_{i-1} - \binom{k_{i-1} + d - i}{d - i + 1}, \quad i \geq 1, \\ m_i &= k_{i-1} - k_i, \quad i \geq 1, \end{aligned}$$

where \mathbb{N}^0 is the set of nonnegative integers; $\binom{n}{k} = 0$ if $k > n$. Define $L = \binom{m+d}{d} = \sum_{k=0}^m \binom{k+d-1}{d-1}$. For $\mathbf{s} = [s_1, \dots, s_d]^\top \in \mathbb{R}^d$, let $\varphi_l(\mathbf{s})$ and ζ_l ($l = 1, \dots, \infty$) be the eigenfunctions and eigenvalues for the modified square exponential kernel $\kappa(\mathbf{s}, \mathbf{s}')$ as defined in (14), then

$$\begin{aligned} \zeta_l &= \left(\frac{\pi}{A}\right)^d B^{k_0}, \quad \frac{\sum_{l=1}^L \zeta_l}{\sum_{l=1}^{\infty} \zeta_l} = (1 - B)^d \sum_{k=0}^m \binom{k+d-1}{d-1} B^k, \\ \varphi_l(\mathbf{s}) &= (2c)^{\frac{d}{4}} \exp(-c\|\mathbf{s}\|_2^2) \prod_{i=1}^d H_{m_i}(\sqrt{2c}s_i), \end{aligned}$$

where $c = \sqrt{a^2 + 2ab}$, $A = a + b + c$ and $B = b/A$; $H_k(\cdot)$ is the k th ($k \in \mathbb{N}^0$) order normalized hermit polynomial, which is defined by $H_k(x) = (2^k k! \sqrt{\pi})^{-1/2} (-1)^k \exp(x^2) \frac{d^k}{dx^k} \exp(-x^2)$.

4.4 A Markov chain Monte Carlo Algorithm

We developed a generally efficient MCMC sampling algorithm for posterior inference about $[\{\tilde{\boldsymbol{\beta}}(\mathbf{s}_i)\}_{i=1}^n, \{\mathbf{u}_k\}_{k=1}^p, \sigma^2, \{\tau_k^2\}_{k=1}^p, \boldsymbol{\lambda} \mid \mathcal{Y}]$ based on the representation and approximation for our model with the TMGP priors (10)-(12).

Updating $\tilde{\boldsymbol{\beta}}(\mathbf{s}_i)$, $i = 1, \dots, n$, is an essential step in the MCMC algorithm. The Metropolis-Hasting (M-H) algorithm is employed with a block updating scheme separately for $\{\tilde{\beta}_k(\mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{R}_g}$, $g = 1, \dots, G$ to facilitate efficient chain mixing. Under the scenario where region partition structure $\mathcal{R}_1, \dots, \mathcal{R}_G$ is available and reliable, we can directly use this partition information. For voxel level analysis or analysis where no prior knowledge about the regional information are adopted, we first fit voxel-wise GLMs and then use certain clustering algorithms to cluster the resulting spatially varying coefficient values. This initial clustering results for the brain locations (usually centers of voxels) are used for block updating. Another M-H step in our algorithm is updating λ_k , $k = 1, \dots, p$, the thresholding parameters in the TMGP priors. The remaining parameters (\mathbf{u}_k , $k = 1, \dots, p$) and hyperparameters ($\sigma^2, \tau_k^2, \kappa = 1, \dots, p$) are

updated by directly drawing samples from their full conditionals due to conjugacy. More details about the MCMC algorithm are available at the appendix.

4.5 Posterior Inference on SVCFs

With the recorded MCMC samples $\tilde{\beta}_k^{(t)}(\mathbf{s}_i)$, \mathbf{u}_k and $\lambda_k^{(t)}$, $t = 1, \dots, T$, we can achieve three major goals: 1) selecting neuroimaging features; 2) estimating covariate effects at the feature regions; 3) making prediction on the covariate effects at any brain location.

To select the important imaging features at the regional level, we estimate the selection probability of every region, $g = 1, \dots, G$, according to the definition (C1)-(C3), using the MCMC samples as

$$\hat{P}(g \in I_1) = \hat{P}\left(\inf_{1 \leq i \leq n, \mathbf{s}_i \in \mathcal{R}_g} |\tilde{\beta}_k(\mathbf{s}_i)| > \lambda_k \mid \mathcal{Y}\right) \approx \frac{1}{T} \sum_{t=1}^T I\left[\inf_{1 \leq i \leq n, \mathbf{s}_i \in \mathcal{R}_g} |\tilde{\beta}_k^{(t)}(\mathbf{s}_i)| > \lambda_k^{(t)}\right],$$

then we estimate $\beta(\mathbf{s}_i)$ as follows, if $\mathbf{s}_i \in \mathcal{R}_g$,

$$\hat{\beta}_k(\mathbf{s}_i) = \begin{cases} \hat{E}[\tilde{\beta}_k(\mathbf{s}_i) \mid \inf_{1 \leq j \leq n, \mathbf{s}_j \in \mathcal{R}_g} |\tilde{\beta}_k(\mathbf{s}_j)| > \lambda_k, \mathcal{Y}], & \hat{P}(g \in I_1) > q \\ 0, & \hat{P}(g \in I_1) \leq q \end{cases}, \quad (15)$$

for all $k = 1, \dots, p$, where $0.5 < q < 1$ is a threshold for the posterior probabilities of being nonzero at certain brain locations. We use $q = 0.90$ throughout the rest of our analysis. Estimates for the posterior conditional expectations in (15) can be easily calculated based on the posterior samples.

As a special case, to conduct voxel level selection (i.e., each voxel is a region with voxel centers being $\mathbf{s}_1, \dots, \mathbf{s}_n$), we can simply adapt (15) to

$$\hat{\beta}_k(\mathbf{s}_i) = \begin{cases} \hat{E}[\tilde{\beta}_k(\mathbf{s}_i) \mid |\tilde{\beta}_k(\mathbf{s}_i)| > \lambda_k, \mathcal{Y}], & \hat{P}(|\tilde{\beta}_k(\mathbf{s}_i)| > \lambda_k \mid \mathcal{Y}) > q \\ 0, & \hat{P}(|\tilde{\beta}_k(\mathbf{s}_i)| > \lambda_k \mid \mathcal{Y}) \leq q \end{cases}, \quad (16)$$

where $\hat{P}(|\tilde{\beta}_k(\mathbf{s}_i)| > \lambda_k \mid \mathcal{Y}) \approx \frac{1}{T} \sum_{t=1}^T I\left[|\tilde{\beta}_k^{(t)}(\mathbf{s}_i)| > \lambda_k^{(t)}\right]$ can be regarded as the posterior probability of activation for each voxel.

Making predication on $\beta(\cdot)$ at an arbitrary new brain location $\mathbf{s}_0 \in \mathcal{R}$ with the posterior samples is also available. Without loss of generality, suppose $\mathbf{s}_0 \in \mathcal{R}_g$, then

$$p(\tilde{\beta}_k(\mathbf{s}_0) \mid \mathcal{Y}) = \int p(\tilde{\beta}_k(\mathbf{s}_0) \mid \tilde{\beta}_k^g, \mathcal{Y}) \times p(\tilde{\beta}_k^g \mid \mathcal{Y}) d\tilde{\beta}_k^g$$

where $p(\cdot)$ represents the probability densities; $\tilde{\beta}_k^g$ is actually $\{\tilde{\beta}_k(\mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{R}_g}$; $[\tilde{\beta}_k(\mathbf{s}_0) \mid \tilde{\beta}_k^g, \mathcal{Y}] = [\tilde{\beta}_k(\mathbf{s}_0) \mid \tilde{\beta}_k^g] \sim N(\boldsymbol{\varphi}(\mathbf{s}_0)^\top \mathbf{u}_k + k_g(\mathbf{s}_0)^\top K_g^{-1}(\tilde{\beta}_k^g - \boldsymbol{\varphi}_g \mathbf{u}_k), \theta^2 - \theta^2 k_g(\mathbf{s}_0)^\top K_g^{-1} k_g(\mathbf{s}_0))$ with

$k_g(\mathbf{s}_0) = \{\kappa(\mathbf{s}_0, \mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{R}_g, 1 \leq i \leq n}$. This indicates that we can predict $\tilde{\beta}_k(\cdot)$ at an unobserved location \mathbf{s}_0 as

$$\hat{\beta}_k(\mathbf{s}_0) = \begin{cases} \frac{1}{T} \sum_{t=1}^T [\boldsymbol{\varphi}(\mathbf{s}_0)^\top \mathbf{u}_k^{(t)} + k_g(\mathbf{s}_0)^\top K_g^{-1}(\tilde{\boldsymbol{\beta}}_k^{g(t)} - \boldsymbol{\varphi}_g \mathbf{u}_k^{(t)})] \tilde{\boldsymbol{\beta}}_k^{g(t)}, & \mathcal{R}_g \text{ is selected} \\ 0, & \mathcal{R}_g \text{ is not selected} \end{cases}, \quad (17)$$

where $\tilde{\boldsymbol{\beta}}_k^{g(t)}$ is the corresponding vector for $\tilde{\boldsymbol{\beta}}_k^g$ based on the t th recorded MCMC sample.

5 Numerical Examples

5.1 Simulation Study: Synthetic 2D Imaging Data

For demonstration purpose, we consider a 2D case, i.e. $\mathcal{R} \subset [0, 1]^2$ in this simulation. three covariate functions, $\beta_1(\mathbf{s})$, $\beta_2(\mathbf{s})$ and $\beta_3(\mathbf{s})$, are created on \mathcal{R} as shown in Figure 3 (the “true” column). We consider $n = 30 \times 30$, 50×50 and 60×60 locations within the fixed domain \mathcal{R} . The data is generated from

$$y_j(\mathbf{s}_i) = \beta_1(\mathbf{s}_i)x_{j1} + \beta_2(\mathbf{s}_i)x_{j2} + \beta_3(\mathbf{s}_i)x_{j3} + e_j(\mathbf{s}_i), \quad (18)$$

where $e_j(\mathbf{s}_i) \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ and $x_{j1} \stackrel{\text{iid}}{\sim} N(0, 4)$, $x_{j2} \stackrel{\text{iid}}{\sim} \text{Unif}(-1, 1)$, $x_{j3} \stackrel{\text{iid}}{\sim} \text{Bernoulli}(0.5)$. We considered two sample sizes $m = 50$ and $m = 100$ in combination with two different noise levels $\sigma^2 = 2$ and $\sigma^2 = 4$. Given one combination $\{n, m, \sigma^2\}$, 50 datasets are independently generated. The MCMC algorithm is implemented to fit the SVCMM model and select informative features for each dataset. We choose the SE kernel for the TMGP priors. The spatial range parameter, b , was fixed as 30. The priors for the thresholding parameters were fixed as $\text{Unif}(0.3, 1.25)$. All pixels are divided into four subsets for block updating according to k -means clustering results. The MCMC iterations are implemented 10,000 times with the first 5,000 samples discarded as burn-in. For each simulated dataset, the algorithm usually takes less than 1 minute to complete the 10,000 iterations on a standard Intel i7 quad core desktop PC.

We compare our results to the standard voxelwise GLM methods. The features estimated from the GLM method are thresholded based on the p -values for testing whether $\beta_k(\mathbf{s}_i) = 0$. We consider both the direct thresholding based on naïve t -test (GLM- t), thresholding using the FDR adjusted p -values (GLM-FDR) (Benjamini and Yekutieli, 2001; Benjamini et al., 2006) and thresholding by controlling FWER based on standard random field theory

(GLM-RFT) (Nichols and Hayasaka, 2003). Figure 3 presents the estimated covariate effect functions from GLM- t , GLM-FDR, GLM-RFT and our method based on one simulated dataset from our experiments. Figure 3 shows that our method provides more accurate feature selection results by eliminating most false positive signals as well as maintaining high sensitivity. Our estimates also recover the features more accurately by incorporating spatial smoothness. In addition, our posterior inference procedures can provide, for each pixel in the images, the probability of presenting features that are different from zero, as shown in the last column in Figure 3. For all the scenarios, we report in Table 1 the relative mean square errors with regard to the GLM estimates, which is defined as

$$\text{ReMSE} = \frac{\sum_{i=1}^n \sum_{k=1}^p \left[\hat{\beta}_k(\mathbf{s}_i) - \beta_k(\mathbf{s}_i) \right]^2}{\sum_{i=1}^n \sum_{k=1}^p \left[\hat{\beta}_k^*(\mathbf{s}_i) - \beta_k(\mathbf{s}_i) \right]^2}, \quad (19)$$

where $\hat{\beta}_k(\mathbf{s}_i)$ are the estimates from a certain method, $\hat{\beta}_k^*(\mathbf{s}_i)$ are the voxel-wise GLM estimates without any thresholding and $\beta_k(\mathbf{s}_i)$ represent the true values. We also reported the false discovery rates (FDRs) and the false negative rates (FNRs) in Table 1, which are specified as:

$$\text{FDR} = \frac{\sum_{i=1}^n \sum_{k=1}^p I[\hat{\beta}_k(\mathbf{s}_i) \neq 0] \times I[\beta_k(\mathbf{s}_i) = 0]}{\sum_{i=1}^n \sum_{k=1}^p I[\hat{\beta}_k(\mathbf{s}_i) \neq 0]}, \quad (20)$$

$$\text{FNR} = \frac{\sum_{i=1}^n \sum_{k=1}^p I[\hat{\beta}_k(\mathbf{s}_i) = 0] \times I[\beta_k(\mathbf{s}_i) \neq 0]}{\sum_{i=1}^n \sum_{k=1}^p I[\beta_k(\mathbf{s}_i) \neq 0]}. \quad (21)$$

Based on the results from Table 1, our method performs well consistently in terms of both feature selection (small FDRs and FNRs) as well as estimation (small ReMSE), especially when the noise level is high or the number of subjects is small. Random field theory based thresholding performs well at low noise level but deteriorates notably as noise level increases due to low sensitivity. FDR control is also relatively robust but this method consistently generates false positive signals. The performance of our SVCM-TMGP method also increases as the number of spatial locations increases within a fixed domain, which agrees with our posterior consistency theory based on infill asymptotics.

To comprehensively compare the performance of TMGP priors in feature selection with other common methods, we also conduct the receiver operating characteristic (ROC) analysis. Since our original method will automatically generate the optimal thresholding values, in this ROC analysis, we fix λ at different values and rerun the MCMC simulation to alternate

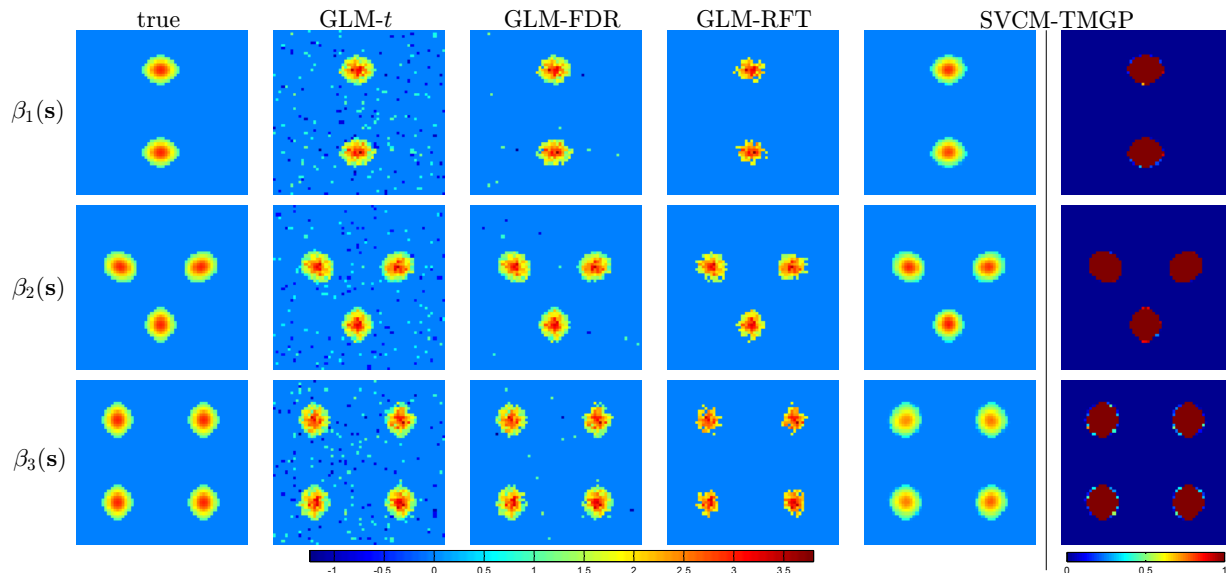


Figure 3: Column 1-5: true and estimated spatial covariate effects from GLM- t , GLM-FDR, GLM-RFT and SVCMTMGP; Column 6: the selection probability estimated from SVCMTMGP. The result is generated from one simulated dataset with $m = 50$ subjects, $n = 3600$ pixels and noise level $\sigma^2 = 4$.

the specificities. Figure 4 shows the ROC curves of our method, GLM-FDR and GLM-RFT with $\sigma^2 = 4$ and $m = 50$ ($n = 900, 2500$ and 3600 , respectively). To quantify the differences, we calculate the area under ROC curves (Table 2) when the false positive rates are smaller than 0.1 for all three methods. Figure 4 and Table 2 show that, under all these three settings, our method (SVCMTMGP) achieve the best performance; as n increases, our method proves to have improved performance. The random field correction based on FWER control (GLM-RFT) suffers from serious false negative problems, leading to low statistical powers. The FDR control (GLM-FDR) is a competitive alternative to our method, especially when the number of spatial locations, n , is relatively small.

To demonstrate the Bayesian learning of the thresholding parameters, we present the histograms for our recorded MCMC samples along with the trace plots for the whole Markov chain from one simulated dataset. It is clear that the marginal posterior distributions of the all thresholding parameters are different from the same prior $\text{Unif}(0.3, 1, 25)$. This indicates our model can achieve Bayesian learning (Xie and Carlin, 2004) of all the thresholding parameters.

Table 1: Quantitative comparison of SVCM-TMGP to voxel-wise GLM fitting results with various thresholdings. All results reported are the means and standard errors based on 50 independently simulated datasets.

		ReMSE	FDR(%)	FNR(%)	ReMSE	FDR(%)	FNR(%)
n=900		$(\sigma^2 = 2, m = 50)$			$(\sigma^2 = 4, m = 50)$		
	GLM- <i>t</i>	0.39(0.12)	18.3(4.3)	0.0(0.0)	0.40(0.23)	17.5(5.3)	1.5(0.2)
	GLM-FDR	0.24(0.07)	5.0(1.0)	0.0(0.0)	0.32(0.08)	5.4(1.5)	4.1(0.6)
	GLM-RFT	0.36(0.17)	0.0(0.0)	4.4(1.5)	0.80(0.29)	0.0(0.0)	22.7(5.6)
	SVCM-TMGP	0.19(0.08)	0.9(0.4)	0.8(0.2)	0.12(0.02)	2.8(0.9)	0.7(0.2)
n=900		$(\sigma^2 = 2, m = 100)$			$(\sigma^2 = 4, m = 100)$		
	GLM- <i>t</i>	0.41(0.13)	18.9(3.5)	0.0(0.0)	0.39(0.11)	19.2(4.9)	0.0(0.0)
	GLM-FDR	0.25(0.06)	5.2(0.7)	0.0(0.0)	0.22(0.06)	5.1(1.2)	0.0(0.0)
	GLM-RFT	0.17(0.10)	0.2(0.0)	0.0(0.0)	0.31(0.14)	0.5(0.0)	4.6(1.3)
	SVCM-TMGP	0.20(0.08)	0.2(0.0)	0.2(0.0)	0.19(0.08)	1.3(0.4)	0.6(0.2)
n=2500		$(\sigma^2 = 2, m = 50)$			$(\sigma^2 = 4, m = 50)$		
	GLM- <i>t</i>	0.32(0.10)	28.1(5.3)	0.0(0.0)	0.33(0.16)	28.8(6.5)	0.4(0.1)
	GLM-FDR	0.13(0.02)	2.9(1.0)	0.3(0.1)	0.20(0.08)	4.3(1.6)	5.0(1.6)
	GLM-RFT	0.27(0.14)	0.3(0.0)	6.3(2.1)	0.61(0.29)	0.4(0.2)	27.9(4.5)
	SVCM-TMGP	0.12(0.06)	0.2(0.0)	0.6(0.1)	0.08(0.03)	1.2(0.4)	0.9(0.4)
n=2500		$(\sigma^2 = 2, m = 100)$			$(\sigma^2 = 4, m = 100)$		
	GLM- <i>t</i>	0.34(0.09)	29.4(4.9)	0.0(0.0)	0.35(0.12)	29.2(5.0)	0.0(0.2)
	GLM-FDR	0.19(0.06)	5.4(0.8)	0.0(0.0)	0.14(0.03)	4.6(1.3)	0.1(0.0)
	GLM-RFT	0.10(0.05)	0.1(0.0)	0.1(0.0)	0.30(0.14)	0.3(0.0)	7.0(2.3)
	SVCM-TMGP	0.15(0.04)	0.1(0.0)	0.3(0.0)	0.13(0.05)	0.4(0.1)	0.6(0.1)
n=3600		$(\sigma^2 = 2, m = 50)$			$(\sigma^2 = 4, m = 50)$		
	GLM- <i>t</i>	0.33(0.09)	35.2(6.2)	0.0(0.0)	0.31(0.11)	31.5(7.3)	0.7(0.4)
	GLM-FDR	0.12(0.03)	3.8(1.0)	0.2(0.0)	0.17(0.04)	3.8(1.3)	5.2(1.4)
	GLM-RFT	0.25(0.07)	0.2(0.0)	7.6(2.2)	0.48(0.10)	0.3(0.1)	28.7(4.6)
	SVCM-TMGP	0.10(0.03)	0.1(0.0)	0.0(0.0)	0.07(0.02)	1.1(0.3)	1.8(0.6)
n=3600		$(\sigma^2 = 2, m = 100)$			$(\sigma^2 = 4, m = 100)$		
	GLM- <i>t</i>	0.33(0.06)	34.4(5.1)	0.0(0.0)	0.33(0.09)	33.5(6.0)	0.0(0.0)
	GLM-FDR	0.18(0.02)	4.3(0.5)	0.0(0.0)	0.13(0.03)	2.6(0.7)	0.0(0.0)
	GLM-RFT	0.09(0.06)	0.0(0.0)	0.2(0.5)	0.24(0.07)	0.1(0.0)	6.5(1.3)
	SVCM-TMGP	0.12(0.03)	0.0(0.0)	0.5(0.0)	0.11(0.03)	0.2(0.0)	0.5(0.3)

Table 2: Area under the ROC curves (AUC) with false positive rates are within $[0, 0.1]$

	AUC $\times 10^{-1}$		
	$n = 900$	$n = 2500$	$n = 3600$
GLM-FDR	0.978 (0.010)	0.976 (0.014)	0.979 (0.005)
GLM-RFT	0.912 (0.049)	0.915 (0.080)	0.911 (0.072)
SVCM-TMGP	0.982 (0.012)	0.989 (0.007)	0.998 (0.001)

5.2 Data Application: The Autism Brain Imaging Data Exchange (ABIDE)

We apply our method to the data from ABIDE, which is a consortium collecting and sharing resting-state fMRI data from 1,112 subjects. Covariate information such as age at scan, sex,

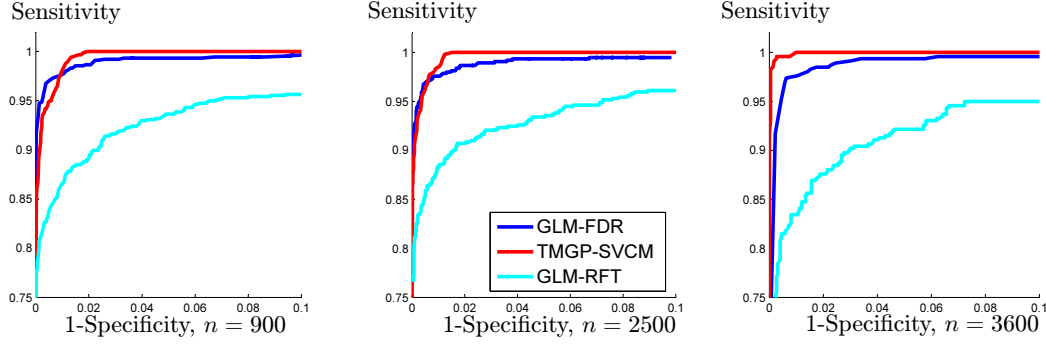


Figure 4: ROC analysis results from 50 replicated datasets. Three competing methods are TMGP-SVCM, GLM-FDR, GLM-RFT. The variance are all $\sigma^2 = 4$; subject numbers are all $m = 50$; the number of spatial locations are $n = 900, 2500$ and 3600 .

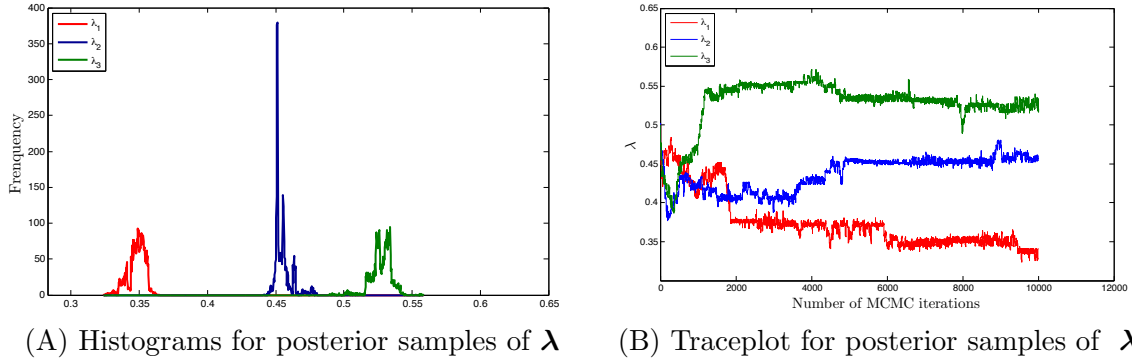


Figure 5: Histograms of λ posterior samples and traceplots for the related Markov chain. Results are generated from one simulation dataset with $n = 3600, \sigma^2 = 4, m = 50$. Initial prior specifications for λ_1, λ_2 and λ_3 are all $\text{Unif}(0.3, 1.25)$.

IQ, handedness and diagnostic information are also available from ABIDE studies. Among the subjects, 539 individuals have Autism spectrum disorders (ASD), which are characterized by symptoms such as social difficulties, communication deficits, stereotyped behaviors and cognitive delays. The remaining subjects are the age-matched normal controls (NC). All the fMRI images are preprocessed through slice-timing, motion correction, nuisance signal regression and temporal filtering. The resulting fMRI data, which are $91 \times 109 \times 91$ 3D matrices, are normalized and registered to Montreal Neurological Institute (MNI) 152 stereotactic space. We aim to investigate the voxel-wise measures of latent functional architecture of

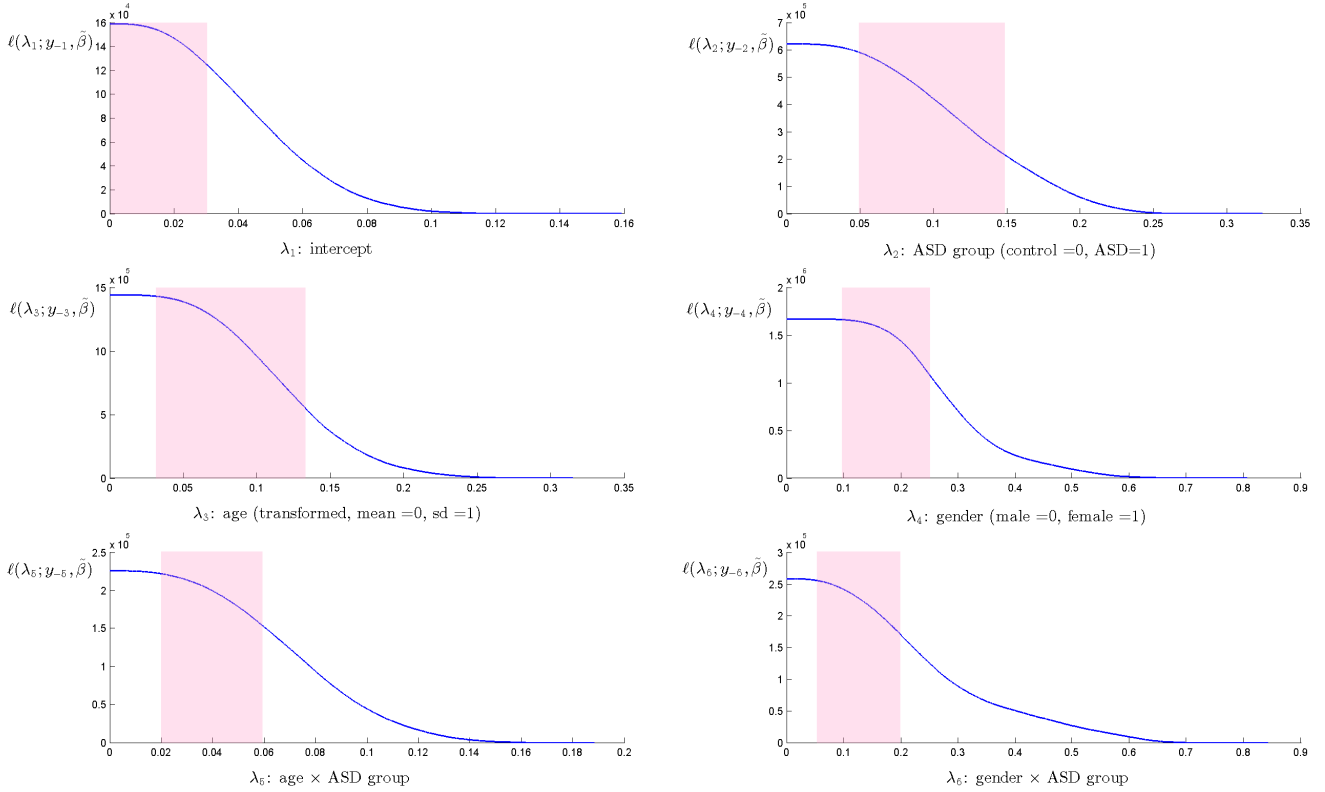


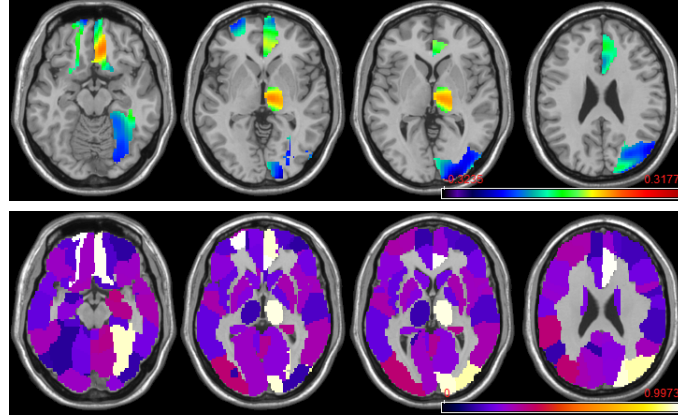
Figure 6: $\ell(\lambda_k)$ for specifying $\boldsymbol{\lambda}$ priors in the analysis of ABIDE data. The colored shades mark the intervals we choose as the range of the uniform priors for $\boldsymbol{\lambda}$.

the brains through fractional amplitude of low-frequency fluctuations (fALFF)(Zou et al., 2008). fALFF is a metric reflecting the percentage of power spectrum within low-frequency domain (0.01 – 0.1Hz) which characterizes the intensity of spontaneous brain activities. We calculate the fALFF for each subject at every voxel. Since the fALFF is restricted to (0, 1), we perform the following monotone transformation

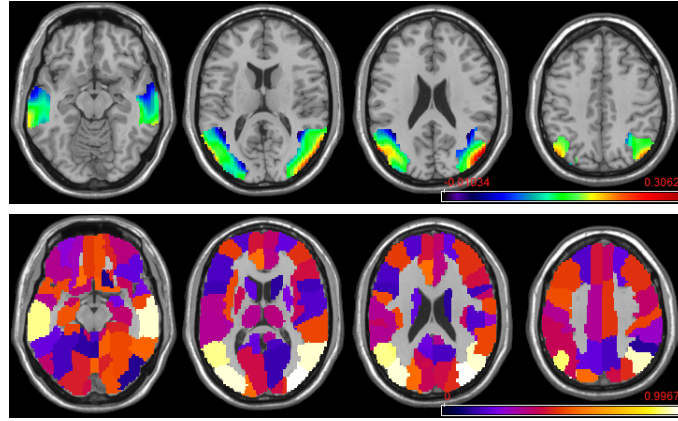
$$y_j(\mathbf{s}) = \log \left(\frac{f_j(\mathbf{s})}{1 - f_j(\mathbf{s})} \right), \quad (22)$$

where $f_j(\mathbf{s})$ represents the fALFF for subject $j = 1, \dots, 1112$ at brain location \mathbf{s} and treat the transformed data as our outcomes.

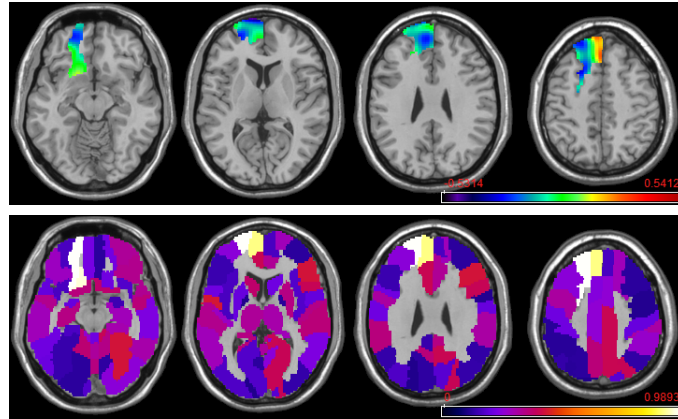
The covariates we choose for fitting model (1) are [1, group, age, gender, group \times age, group \times gender]. We use all the voxels at the gray matter as the observed spatial locations $\mathbf{s}_1, \dots, \mathbf{s}_n$ and all the anatomical parcellation based on MNI templates as our brain regions



(A) Covariate effects for the ASD group versus the control



(B) Covariate effects for the age



(C) Covariate effects for group and gender interaction

Figure 7: Estimated SVCFs (top row in each subplot) and regional selection probabilities (bottom row in each subplot) based on posterior samples from our MCMC algorithm for “ASD group”, “age” and “ASD group \times gender”

$\mathcal{R}_1, \dots, \mathcal{R}_G$ ($n = 177,743$ and $G = 116$). The imaging outcomes are centered across all subjects at each voxel. The group variable equals to 1 for the ASD subjects; the ages are all centered and scaled with zero mean and unit variance; the gender variable equals to 1 for female subjects. The priors for the thresholding parameters are determined through the method described in subsection 4.2. The profiles of $\hat{\ell}(\lambda_k)$ are shown in Figure 6. The priors for the six thresholding parameters are $\text{Unif}(0, 0.03)$, $\text{Unif}(0.05, 0.15)$, $\text{Unif}(0.03, 0.13)$, $\text{Unif}(0.1, 0.25)$, $\text{Unif}(0.02, 0.06)$ and $\text{Unif}(0.05, 0.2)$ according to the plots. The Gaussian kernel we use is the modified square exponential kernel with $a = 0.25, b = 95$ (bandwidth fixed without updating). To achieve 90% recovery rate of the KL expansion, we set $L = 1,140$ eigenfunctions. The MCMC algorithm runs 60,000 iterations with 25,000 burn-in.

Based on our results, the ASD subjects tend to show lower fALFF outcomes at the median and superior part of the right occipital lobe, which is the visual processing centers of human brains (the visual cortex). We observe significantly higher activities at the right fusiform gyrus, which has been reported to be related to Autism in (Hadjikhani et al., 2004). Similar findings are observed at the right median orbitofrontal cortex, the region involved in most human cognition processes, especially decision-making, indicating more spontaneous brain cognition activities among the ASD subjects. From the axial view, Figure 7(A) shows the information discussed above. Some other regions that are selected includes the right thalamus and the right anterior cingulum, which we do not discuss here in detail.

Another major findings are the age effect on the fALFFs. We identified three brain regions that show higher fALFF outcomes as the age increases: the median occipital lobe, the median temporal lobe and the angular gyrus. These regions are generally involved in brain functions such as spatial temporal cognition, language, memory, attention and visual processing. Figure 7(B) shows the findings above in brain slices from the axial view.

Although no specific regions of interest are observed for the “gender” variable, certain brain regions demonstrate different activation patterns its interaction with the disease group. Specifically, female ASD subjects have higher fALFFs as compared with male ASD effects at the left median and superior part of the orbital gyrus but lower fALFFs at the left frontal lobe gyrus and the left rectus. Figure 7(C) shows these findings in three views for the ease of demonstration. Beyond these findings, we also note that the right inferior temporal gyrus

displays smaller effects among the female ASD subjects as compared with the male autistics.

6 Discussion

In this paper, we introduced a new family of prior, the TMGP prior, for feature selections within spatially varying coefficient functions and its applications to massive neuroimaging data analysis. We demonstrate the prior large support properties of the TMGP prior and its posterior consistency under the spatially varying coefficient models under the spatial infill asymptotics. Simulation studies show that the TMGP prior is especially useful for imaging feature selections with relatively large noise or small sample sizes.

In most spatial statistics literatures such as [Diggle et al. \(1998\)](#); [Gelfand et al. \(2003\)](#); [Smith et al. \(2002\)](#), a spatial process are decomposed into three parts: a deterministic trend process or, in other words, mean process; a zero-mean variance process with continuous sample path and a zero-mean white noise process, i.e., the nugget effect. Under this general framework, [Zhu et al. \(2014\)](#) considered a more complex model as compared with our model (1), which could be expressed using our notations as

$$y_j(\mathbf{s}) = \mathbf{x}_j^T \boldsymbol{\beta}(\mathbf{s}) + \eta_j(\mathbf{s}) + e_j(\mathbf{s}), \quad (23)$$

for all subjects $j = 1, \dots, m$. They estimated the additional term $\eta_j(\mathbf{s})$ for every subject through standard local linear regression techniques, which do not require a pre-specified spatial covariance structure $\boldsymbol{\Sigma}_\eta$, or equivalently, its Karhunen-Loève basis. Since in this paper, our primary focus is on the GLM framework for imaging data, we did not apply our TMGP prior under the setting of model (23). To enable a similar analysis using the TMGP prior for $\boldsymbol{\beta}(\mathbf{s})$ in (23) under the Bayesian framework, there are four major tasks. First, a proper prior specification for $\eta_j(\mathbf{s})$ needs to be introduced which should be flexible enough to capture various spatially smooth dynamics. Second, a computationally efficient algorithm for estimating the additional parameters brought by $\eta_j(\mathbf{s})$ is required since this set of parameters scale with the number of subjects. Third, since for each subject, we will have a subject specific random effect term, we need to carefully monitor the model fitting procedures to avoid potential over-fitting issues. Fourth, the theoretical analysis for posterior consistency in Theorem 2 needs to be adapted to the more challenging model structure.

In addition to applying the TMGP prior to model (23), our study can be extended to some other directions. With a focus on neuroimaging studies using model (1), we can extend

the prior constructions to enable self-guided parcellation while conducting feature selection tasks. This can help relax the region based sparsity assumptions for the SVCFs. We can also explore the Bayesian asymptotic theories when the number of brain region partitions diverges. With a focus on general Bayesian analysis, we can extend TMGP for modeling high-dimensional multivariate binary processes or selecting features for scalar-on-image models such as neuropsychiatric disease predictions.

Appendices

A Proof of Theorem 1

Based on the assumptions for $\beta^0(\mathbf{s})$, let $I_1 = I_1[\beta^0(\mathbf{s})]$ and $I_0 = \{1, \dots, G\} \setminus I_1$, we have that

$$\Pi(\|\beta(\mathbf{s}) - \beta^0(\mathbf{s})\|_\infty < \varepsilon) \geq \Pi\left(\sup_{\mathbf{s} \in \cup_{g \in I_1} \mathcal{R}_g} |\tilde{\beta}(\mathbf{s}) - \beta^0(\mathbf{s})| < \varepsilon, \inf_{\mathbf{s} \in \cup_{g \in I_1} \mathcal{R}_g} |\tilde{\beta}(\mathbf{s})| > \lambda, \sup_{\mathbf{s} \in \cup_{g \in I_0} \mathcal{R}_g} |\tilde{\beta}(\mathbf{s})| \leq \lambda\right). \quad (24)$$

Without loss of generality, we only consider $0 < \varepsilon < \lambda_0 - \lambda$. Note that for all $\mathbf{s} \in \cup_{g \in I_1} \mathcal{R}_g$, $|\tilde{\beta}(\mathbf{s}) - \beta^0(\mathbf{s})| < \varepsilon$ and $|\beta^0(\mathbf{s})| \geq \lambda_0$ implies that $|\tilde{\beta}(\mathbf{s})| \geq \lambda_0 - \varepsilon > \lambda$, then (24) is equivalent to

$$\Pi(\|\beta(\mathbf{s}) - \beta^0(\mathbf{s})\|_\infty < \varepsilon) \geq \Pi\left(\sup_{\mathbf{s} \in \cup_{g \in I_1} \mathcal{R}_g} |\tilde{\beta}(\mathbf{s}) - \beta^0(\mathbf{s})| < \varepsilon, \sup_{\mathbf{s} \in \cup_{g \in I_0} \mathcal{R}_g} |\tilde{\beta}(\mathbf{s})| \leq \lambda\right).$$

Let $\phi_l(\mathbf{s})$ and $\zeta_l, l = 1, \dots, \infty$, be the normalized eigenfunctions and eigenvalues of the kernel function $\kappa(\cdot, \cdot)$, then the Karhunen-Loève expansions of $\gamma(\mathbf{s})$ and $\epsilon(\mathbf{s})$ can be expressed as $\gamma(\mathbf{s}) = \sum_{l=1}^\infty u_l \phi_l(\mathbf{s}), \forall \mathbf{s} \in \mathcal{R}$, and $\epsilon(\mathbf{s}) = \sum_{l=1}^\infty v_{lg} \phi_l(\mathbf{s}), \mathbf{s} \in \mathcal{R}_g, g = 1, \dots, G$, such that $u_l \stackrel{\text{iid}}{\sim} N(0, \zeta_l \tau^2), v_{lg} \stackrel{\text{iid}}{\sim} N(0, \zeta_l \theta^2)$ and u_l, v_{lg} are all independent. Since the RKHS of $\kappa(\cdot, \cdot)$ is $\mathcal{C}(\mathcal{R})$, for \mathbf{s} within any partition \mathcal{R}_g , $\beta^0(\mathbf{s})$ can be represented as $\sum_{l=1}^\infty w_{lg} \phi_l(\mathbf{s})$, where $\sum_{l=1}^\infty w_{lg}^2 < \infty$.

For $\mathbf{s} \in \mathcal{R}_g$ with $g \in I_1$

$$\sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}(\mathbf{s}) - \beta^0(\mathbf{s})| \leq \sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}(\mathbf{s}) - \beta_{L,g}^0(\mathbf{s})| + \sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}^*(\mathbf{s})| + \sup_{\mathbf{s} \in \mathcal{R}_g} |\beta_{L,g}^{0*}(\mathbf{s})|, \quad (25)$$

where $\tilde{\beta}_{L,g}(\mathbf{s}) = \sum_{l=1}^L (u_l + v_{lg}) \phi_l(\mathbf{s}), \beta_{L,g}^0(\mathbf{s}) = \sum_{l=1}^L w_{lg} \phi_l(\mathbf{s}), \tilde{\beta}_{L,g}^*(\mathbf{s}) = \tilde{\beta}(\mathbf{s}) - \tilde{\beta}_{L,g}(\mathbf{s})$

and $\beta_{L,g}^{0*}(\mathbf{s}) = \beta^0(\mathbf{s}) - \beta_{L,g}^0(\mathbf{s})$. Since the RKHS of $\kappa(\cdot, \cdot)$ is $\mathcal{C}(\mathcal{R})$, $\tilde{\beta}(\mathbf{s})$ is uniformly continuous on \mathcal{R}_g with probability 1, then by Theorem 3.1.2 of [Adler and Taylor \(2009\)](#), $\lim_{L \rightarrow \infty} \sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}^*(\mathbf{s})| = 0$ with probability 1. By the uniform convergence of the series $\sum_{l=1}^L w_{lg} \phi_l(\mathbf{s})$ to $\beta^0(\mathbf{s})$ as $L \rightarrow \infty$ on \mathcal{R}_g , $\lim_{L \rightarrow \infty} \sup_{\mathbf{s} \in \mathcal{R}_g} |\beta_{L,g}^{0*}(\mathbf{s})| = 0$. Then we can find a finite integer L_g such that for all $L \geq L_g$, $\sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}^*(\mathbf{s})| < \frac{\varepsilon}{3}$ with probability 1 and $\sup_{\mathbf{s} \in \mathcal{R}_g} |\beta_{L,g}^{0*}(\mathbf{s})| < \frac{\varepsilon}{3}$. Since $\phi_l(\mathbf{s}), l = 1, \dots, L_g$ are all continuous functions in on \mathcal{R} , we have that $\max_{1 \leq l \leq L_g} \|\phi_l(\mathbf{s})\|_\infty < M_{\phi, L_g}$ where M_{ϕ, L_g} is a certain constant. Let $|u_l + v_{lg} - w_{lg}| < \frac{\varepsilon}{3L_g M_{\phi, L_g}}$ for all $l = 1, \dots, L_g$ and consider $L = L_g$ in (25), we have that $\sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}(\mathbf{s}) - \beta_{L,g}^0(\mathbf{s})| < \frac{\varepsilon}{3}$. Thus, the condition $|u_l + v_{lg} - w_{lg}| < \frac{\varepsilon}{3L_g M_{\phi, L_g}}, l = 1, \dots, L_g$ can guarantee that $\sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}(\mathbf{s}) - \beta^0(\mathbf{s})| < \varepsilon$ with probability 1 for $g \in I_1$.

For $\mathbf{s} \in \mathcal{R}_g$ with $g \in I_0$, similar to (25) and the definitions above, we have

$$\sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}(\mathbf{s})| \leq \sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}(\mathbf{s})| + \sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}_{L,g}^*(\mathbf{s})|. \quad (26)$$

Similarly, we can find L_g and M_{ϕ, L_g} such that $|u_l + v_{lg}| \leq \frac{\lambda}{2L_g M_{\phi, L_g}}, l = 1, \dots, L_g$ guarantees that $\sup_{\mathbf{s} \in \mathcal{R}_g} |\tilde{\beta}(\mathbf{s})| \leq \lambda$ with probability 1 for all $g \in I_0$.

Then we have

$$\begin{aligned} \Pi(\|\beta(\mathbf{s}) - \beta^0(\mathbf{s})\|_\infty < \varepsilon) &\geq \Pi\left(\left\{|u_l + v_{lg} - w_{lg}| < \frac{\varepsilon}{3L_g M_{\phi, L_g}} : l = 1, \dots, L_g, g \in I_1\right\} \cup \right. \\ &\quad \left.\left\{|u_l + v_{lg}| \leq \frac{\lambda}{2L_g M_{\phi, L_g}} : l = 1, \dots, L_g, g \in I_0\right\}\right) > 0, \end{aligned} \quad (27)$$

due to the positive measures assigned on arbitrary nonempty sets by the $\left(\sum_{g=1}^G L_g + L_{\max}\right)$ -dimensional multivariate Gaussian distribution: $(u_1, \dots, u_{L_{\max}}, v_{11}, \dots, v_{L_1 1}, \dots, v_{1G}, \dots, v_{L_G G})$, where $L_{\max} = \max_{g=1, \dots, G} L_g$.

B Proof of Theorem 2

B.1 KL neighborhood conditions for noniid outcomes

Lemma 1. Consider our observed data $\mathbf{y}(\mathbf{s}_i) \in \mathbb{R}^m$, $\mathbf{y}(\mathbf{s}_i) \sim f_{i, \beta}(\mathbf{y})$ where

$$f_{i, \beta}(\mathbf{y}) = (2\pi\sigma^2)^{-m/2} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{X}\beta(\mathbf{s}_i)\|_2^2\right\}$$

for some constant $\sigma^2 > 0$. Define

$$D_i(\beta_0, \beta) = \log \frac{f_{i, \beta_0}}{f_{i, \beta}}, \quad K_i(\beta_0, \beta) = E_{f_{i, \beta_0}}[D_i(\beta_0, \beta)], \quad V_i(\beta_0, \beta) = \text{Var}_{f_{i, \beta_0}}[D_i(\beta_0, \beta)].$$

If we assign an independent TMGP prior for each dimension of β , i.e.,

$$\beta_k(\mathbf{s}) \sim \mathcal{TMGP}(\tau_k^2, \theta_k^2, \lambda_k, \kappa(\cdot, \cdot)),$$

then we have that $\exists B, \Pi(B) > 0$ such that

$$\liminf_{n \rightarrow \infty} \Pi \left(\left\{ \beta \in B : \frac{1}{n} \sum_{i=1}^n K_i(\beta^0, \beta) < \varepsilon \right\} \right) > 0,$$

$$\frac{1}{n^2} \sum_{i=1}^n V_i(\beta^0, \beta) \rightarrow 0, \forall \beta \in B.$$

Proof. It is trivial to have that

$$D_i(\beta_0, \beta) = \frac{1}{\sigma^2} (\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i))^\top \left[\mathbf{X}^\top \mathbf{y} - \frac{1}{2} \mathbf{X} \mathbf{X}^\top (\beta^0(\mathbf{s}_i) + \beta(\mathbf{s}_i)) \right],$$

then since $E_{f_{i, \beta_0}}[\mathbf{y}] = \mathbf{X} \beta(\mathbf{s}_i)$, $Var_{f_{i, \beta_0}}[\mathbf{y}] = \sigma^2 \mathbf{I}_m$,

$$K_i(\beta_0, \beta) = \frac{1}{2\sigma^2} (\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i))^\top \mathbf{X}^\top \mathbf{X} (\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i)) \leq \frac{md_{\max}}{2\sigma^2} \|\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i)\|_2^2,$$

$$V_i(\beta_0, \beta) = \frac{1}{\sigma^2} (\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i))^\top \mathbf{X}^\top \mathbf{X} (\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i)) \leq \frac{md_{\max}}{\sigma^2} \|\beta^0(\mathbf{s}_i) - \beta(\mathbf{s}_i)\|_2^2.$$

Now consider $B_k = \left\{ \beta_k(\mathbf{s}) : \|\beta_k(\mathbf{s}) - \beta^0(\mathbf{s})\|_\infty < \sqrt{\frac{2\sigma^2\varepsilon}{mpd_{\max}}} \right\}$ and let $B = \cap_{k=1}^p B_k$. Since the priors for $\beta_k(\mathbf{s}), k = 1, \dots, p$ are independent and $\Pi(B_k) > 0$ due to Theorem (1), $\Pi(B) = \prod_{k=1}^p \Pi(B_k) > 0$.

For all $\beta \in B$, $K_i(\beta_0, \beta) \leq \frac{md_{\max}}{2\sigma^2} \sum_{k=1}^p \|\beta_k^0(\mathbf{s}) - \beta_k(\mathbf{s})\|_2^2 < \varepsilon$ and similarly $V_i(\beta_0, \beta) < 2\varepsilon$. Then $\liminf_{n \rightarrow \infty} \Pi \left(\left\{ \beta \in B : \frac{1}{n} \sum_{i=1}^n K_i(\beta^0, \beta) < \varepsilon \right\} \right) = \Pi(B) > 0$ and $\frac{1}{n^2} \sum_{i=1}^n V_i(\beta^0, \beta) < \frac{2\varepsilon}{n} \rightarrow 0$ for all $\beta \in B$. \square

B.2 Sieve constructions

Define the set of functions

$$\mathcal{P}_n = \left\{ \beta(\mathbf{s}) \in \mathcal{P} : \|\beta(\mathbf{s})\|_\infty < \sqrt{n}, \sup_{\mathbf{s} \in \mathcal{R}_g} |D^\alpha \beta(\mathbf{s})| < \sqrt{n}, g \in I_1[\beta(\mathbf{s})], 1 \leq \|\alpha\|_1 \leq \rho \right\},$$

as our sieve construction.

Lemma 2. *If $G < \infty$, the ε -covering number under the sup-norm for \mathcal{P}_n satisfies*

$$\log N(\varepsilon, \mathcal{P}_n, \|\cdot\|_\infty) < C n^{\frac{d}{2\rho}} \varepsilon^{-d},$$

for some finite constant C .

Proof. Define

$$\mathcal{P}_{n,g} = \left\{ \beta(\mathbf{s}) \in \mathcal{C}^\rho(\overline{\mathcal{R}_g}) : \sup_{\mathbf{s} \in \mathcal{R}_g} |D^\alpha \beta(\mathbf{s})| < \sqrt{n}, 0 \leq \|\alpha\|_1 \leq \rho \right\},$$

for all $g = 1, \dots, G$. Theorem 2.7.1 of [Van Der Vaart and Wellner \(1996\)](#) implies that

$$\log N(\varepsilon, \mathcal{P}_{n,g}, \|\cdot\|_\infty) \leq C_g n^{\frac{d}{2\rho}} \varepsilon^{-d},$$

for some constants $C_g < \infty$. Then by the definition of \mathcal{P}_n , we have that

$$N(\varepsilon, \mathcal{P}_n, \|\cdot\|_\infty) \leq \prod_{g=1}^G N(\varepsilon, \mathcal{P}_{n,g}, \|\cdot\|_\infty) \leq \exp \left\{ C n^{\frac{d}{2\rho}} \varepsilon^{-d} \right\},$$

where $C = \sum_{g=1}^G C_g < \infty$. \square

Lemma 3. *Consider the TMGP prior for $\beta(\mathbf{s})$ with kernel function satisfying condition (K1)(K2), then $\Pi(\mathcal{P} \cap \mathcal{P}_n^c) \leq D e^{-bn}$ for some constant $D, b > 0$.*

Proof. The construction of the TMGP prior implies that

$$\Pi(\mathcal{P}_n) \geq \prod_{g=1}^G \Pi \left(\sup_{\mathbf{s} \in \mathcal{R}_g} |D^\alpha \tilde{\beta}_g(\mathbf{s})| > \sqrt{n}, 0 \leq \|\alpha\|_1 \leq \rho \right),$$

where $\tilde{\beta}_g(\mathbf{s}) \stackrel{\text{iid}}{\sim} \mathcal{GP}(0, (\theta^2 + \tau^2) \kappa(\mathbf{s}, \mathbf{s}'))$ for all $g = 1, \dots, G$. By applying Theorem 5 of Ghosal et al. (2006), we have that $\Pi \left(\sup_{\mathbf{s} \in \mathcal{R}_g} |D^\alpha \tilde{\beta}_g(\mathbf{s})| > \sqrt{n}, 0 \leq \|\alpha\|_1 \leq \rho \right) \geq 1 - A e^{-bn}$ for some $A, b > 0$, given that $\kappa(\mathbf{s}, \cdot)$ has continuous partial derivatives of order $2\rho + 2$ on the compact set \mathcal{R} . Then we have that $\Pi(\mathcal{P} \cap \mathcal{P}_n^c) \leq 1 - (1 - A e^{-bn})^G \leq D e^{-dn}$ where $D = AG$ due to the fact that $(1 - x)^G \geq 1 - Gx$ for all $0 < x < 1$ and $G = 1, 2, \dots$. \square

Now we define $\mathbf{P}_n = \{\beta(\mathbf{s}) = [\beta_1(\mathbf{s}), \dots, \beta_p(\mathbf{s})]^\top : \beta_k(\mathbf{s}) \in \mathcal{P}_n, k = 1, \dots, p\}$ here and below. Then we can easily get that

$$N(\varepsilon, \mathbf{P}_n, \|\cdot\|_\infty) < \exp\{C p n^{\frac{d}{2\rho}} \varepsilon^{-d}\}, \quad (28)$$

and if assign TMGP priors independently for all elements in $\beta(\mathbf{s})$ then

$$\Pi(\mathbf{P} \cap \mathbf{P}_n^c) \leq D p \exp\{-bn\}. \quad (29)$$

B.3 Test Constructions

Lemma 4. *Consider $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 \mathbf{I}_m)$, a standard linear model with sample size m where $\mathbf{y} = [y_1, \dots, y_m]$; \mathbf{X} is an $m \times p$ design matrix satisfying assumption (X1). Consider the test function $\Phi = I \left(\|\hat{\beta} - \beta^0\|_2 > \frac{\varepsilon \sqrt{p}}{2} \right)$ for testing $H_0 : \beta = \beta^0$ versus $H_1 : \beta = \beta^1$, where $\beta^0 \in \mathbb{R}^p$ and $\beta^1 \in \{\beta \in \mathbb{R}^p : \|\beta - \beta^0\|_\infty \geq \varepsilon\}$; $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is the ordinary least square estimator. Then for $m > \frac{8(\log 2)\sigma^2}{\varepsilon^2 d_{\min}}$, we have that*

$$E_{P_0}[\Phi] \leq \exp\{-\Omega_m m p\}, \quad E_{P_1}[1 - \Phi] \leq \exp\{-\Omega_m m p\},$$

for some $\Omega_m > 0$ depending on m , where P_0 and P_1 represents the probability distributions under H_0 and H_1 .

Proof. Note that for $t = 0, 1$, under H_t , $\|\hat{\beta} - \beta^t\|_2^2 d_{\min} m / \sigma^2 \leq (\hat{\beta} - \beta^t)^\top \mathbf{X}^\top \mathbf{X} (\hat{\beta} - \beta^t) / \sigma^2 \sim \chi_p^2$, then we have that

$$E_{P_0}[\Phi] = P_0 \left(\|\hat{\beta} - \beta_0\|_2^2 > \frac{\varepsilon^2 p}{4} \right) \leq P_0 \left(\chi_p^2 > \frac{\varepsilon^2 p d_{\min} m}{4\sigma^2} \right) \leq \exp \left\{ - \left(\frac{\varepsilon^2 d_{\min}}{16\sigma^2} - \frac{\log 2}{2m} \right) mp \right\}, \quad (30)$$

where the last inequality is simply due to the fact that $P(\chi_p^2 > x) \leq (1-2t)^{-p/2} \exp\{-tx\}$, $\forall 0 < t < 1/2$ by letting $t = 1/4$. Similarly,

$$\begin{aligned} E_{P_1}[1 - \Phi] &= P_1 \left(\|\hat{\beta} - \beta^0\|_2 \leq \frac{\varepsilon \sqrt{p}}{2} \right) \\ &\leq P_1 \left(\left| \|\hat{\beta} - \beta^1\|_2 - \|\beta^0 - \beta^1\|_2 \right| \leq \frac{\varepsilon \sqrt{p}}{2} \right) \\ &\leq P_1 \left(\|\hat{\beta} - \beta^1\|_2 \geq -\frac{\varepsilon \sqrt{p}}{2} + \|\beta^0 - \beta^1\|_2 \right) \\ &\leq P_1 \left(\|\hat{\beta} - \beta^1\|_2 \geq \frac{\varepsilon \sqrt{p}}{2} \right) \\ &\leq \exp \left\{ - \left(\frac{\varepsilon^2 d_{\min}}{16\sigma^2} - \frac{\log 2}{2m} \right) mp \right\}. \end{aligned} \quad (31)$$

Define $\Omega_m = \frac{\varepsilon^2 d_{\min}}{16\sigma^2} - \frac{\log 2}{2m}$ here and below. Notice that $m > \frac{8(\log 2)\sigma^2}{\varepsilon^2 d_{\min}}$ is equivalent to $\Omega_m > 0$, we complete the proof. \square

Lemma 5. *We consider n_0 locations $\mathbf{s}_1, \dots, \mathbf{s}_{n_0}$ and define $\beta_i = [\beta_1(\mathbf{s}_i), \dots, \beta_p(\mathbf{s}_i)]^\top$ (β_i^0 and β_i^1 can be defined accordingly). Suppose that we have observed the data $\mathbf{y}_i = [y_1(\mathbf{s}_i), \dots, y_m(\mathbf{s}_i)]^\top$ generated from the SVCMM, i.e. $\mathbf{y}_i \sim N(\mathbf{X}\beta_i, \sigma^2 \mathbf{I}_m)$. Consider testing $H_0 : \beta_i = \beta_i^0, i = 1, \dots, n_0$ versus $H_1 : \beta_i = \beta_i^1, i = 1, \dots, n_0$ where $\|\beta_i^1 - \beta_i^0\|_\infty \geq \varepsilon$ for all $i = 1, \dots, n_0$. Define $\Phi_i = I \left(\|\hat{\beta}_i - \beta_i^0\|_2 > \frac{\varepsilon \sqrt{p}}{2} \right)$ with $\hat{\beta}_i = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}_i$. Then for the test function $\tilde{\Phi} = I \left(\sum_{i=1}^{n_0} \Phi_i > \frac{n_0}{2} \right)$, we have that*

$$E_{P_0}[\tilde{\Phi}] \leq \exp\{-Cn_0\}, \quad E_{P_1}[1 - \tilde{\Phi}] \leq \exp\{-Cn_0\}$$

Proof. By the results from Lemma 4, $E_{P_0}[\Phi_i] \leq e^{-\Omega_m mp}$ and $E_{P_1}[1 - \Phi_i] \leq e^{-\Omega_m mp}$ for all $i = 1, \dots, n_0$. Then

$$E_{P_0}[\tilde{\Phi}] \leq P_0 \left(\sum_{i=1}^{n_0} \Phi_i - \sum_{i=1}^{n_0} E_{P_0}[\Phi_i] > \frac{n_0}{2} - n_0 e^{-\Omega_m mp} \right), \quad (32)$$

$$E_{P_1}[1 - \tilde{\Phi}] = P_1 \left(\sum_{i=1}^{n_0} (1 - \Phi_i) \geq \frac{n_0}{2} \right) \leq P_1 \left(\sum_{i=1}^{n_0} (1 - \Phi_i) - \sum_{i=1}^{n_0} E[1 - \Phi_i] \geq \frac{n_0}{2} - n_0 e^{-\Omega_m mp} \right). \quad (33)$$

By the Hoeffding inequality (Hoeffding, 1963), the right hand side of both (32) and (33) are bounded by $\exp\left\{-\frac{2}{n_0} \frac{n_0^2(1-2e^{-\Omega_m mp})^2}{4}\right\}$ when $1 - 2e^{-\Omega_m mp} > 0$. That is, $E_{P_0}[\tilde{\Phi}] \leq \exp\{-Cn_0\}$, $E_{P_1}[1 - \tilde{\Phi}] \leq \exp\{-Cn_0\}$ where $C = \frac{(1-2e^{-\Omega_m mp})^2}{2}$. \square

Lemma 6. For two functions $\beta^0(\mathbf{s}), \beta^1(\mathbf{s}) \in \mathcal{P}$, if $\|\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})\|_1 = \int_{\mathbf{s} \in \mathcal{R}} |\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \mathbb{P}_n(d\mathbf{s}) \geq \varepsilon$, we have that $\mathbb{P}_n(|\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \geq \frac{\varepsilon}{2}) \geq c'$ where $0 < c' \leq 1$ is a constant. That is, the set $\{\mathbf{s} \in \{\mathbf{s}_1, \dots, \mathbf{s}_n\} : |\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \geq \frac{\varepsilon}{2}\}$ has $n_0 \geq c'n - 1$ elements.

Proof. Let $\mathcal{S} = \{\mathbf{s} \in \mathcal{R} : |\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \geq \frac{\varepsilon}{2}\}$, then

$$\begin{aligned} \varepsilon &\leq \int_{\mathbf{s} \in \mathcal{R}} |\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \mathbb{P}_n(d\mathbf{s}) \\ &= \int_{\mathbf{s} \in \mathcal{S}} |\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \mathbb{P}_n(d\mathbf{s}) + \int_{\mathbf{s} \in \mathcal{R} \setminus \mathcal{S}} |\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \mathbb{P}_n(d\mathbf{s}) \\ &\leq (M_0 + M_1) \mathbb{P}_n\left(|\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \geq \frac{\varepsilon}{2}\right) + \frac{\varepsilon}{2} \mathbb{P}_n(\mathcal{R}), \end{aligned}$$

where $\mathbb{P}_n(\mathcal{R}) = 1$; $M_0 = \|\beta^0(\mathbf{s})\|_\infty$ and $M_1 = \|\beta^1(\mathbf{s})\|_\infty$ are finite constants due to absolute continuity. Thus $\mathbb{P}_n(|\beta^0(\mathbf{s}) - \beta^1(\mathbf{s})| \geq \frac{\varepsilon}{2}) \geq c'$ by letting $c' = \frac{\varepsilon}{2(M_0 + M_1)}$. \square

Lemma 7. There exists a test Φ_{β^1, β^0} for testing $H_0 : \beta(\mathbf{s}) = \beta^0(\mathbf{s})$ against $H_1 : \beta(\mathbf{s}) = \beta^1(\mathbf{s})$ where $\|\beta^1(\mathbf{s}) - \beta^0(\mathbf{s})\|_{1, \infty} \geq \varepsilon$ in our proposed SVCN, such that

$$E_{P_0}[\Phi_{\beta^1, \beta^0}] \leq \exp\{-Cn\}, \quad E_{P_1}[1 - \Phi_{\beta^1, \beta^0}] \leq \exp\{-Cn\},$$

for some constant C with P_0 and P_1 corresponding to the probability distributions under H_0 and H_1 .

Proof. For two vector-valued functions $\beta^t(\mathbf{s}) = [\beta_1^t(\mathbf{s}), \dots, \beta_p^t(\mathbf{s})]^\top, t = 0, 1$, if $\|\beta^1(\mathbf{s}) - \beta^0(\mathbf{s})\|_{1, \infty} \geq \varepsilon$, we must have at least one $k \in \{1, \dots, p\}$, such that $\|\beta_k^1(\mathbf{s}) - \beta_k^0(\mathbf{s})\|_1 \geq \varepsilon$, then due to Lemma 6, we can find $n_0 \geq c'n - 1$ elements in $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ such that $|\beta_k^1(\mathbf{s}) - \beta_k^0(\mathbf{s})| \geq \frac{\varepsilon}{2}$. Without loss of generality, we denote these points as $\mathbf{s}_1, \dots, \mathbf{s}_{n_0}$. Then for all $\mathbf{s}_i, i = 1, \dots, n_0$, we have that $\|\beta^1(\mathbf{s}_i) - \beta^0(\mathbf{s}_i)\|_\infty \geq \frac{\varepsilon}{2}$.

Now define the set $\mathcal{S}_{\beta^1, \beta^0} = \{\mathbf{s} \in \{\mathbf{s}_1, \dots, \mathbf{s}_n\} : \|\beta^1(\mathbf{s}_i) - \beta^0(\mathbf{s}_i)\|_\infty \geq \frac{\varepsilon}{2}\}$. Then $n_0 = |\mathcal{S}_{\beta^1, \beta^0}| \geq c'n - 1$. Define the test function

$$\Phi_{\beta^1, \beta^0} = I\left(\sum_{\mathbf{s} \in \mathcal{S}_{\beta^1, \beta^0}} \Phi(\mathbf{s}) > \frac{n_0}{2}\right),$$

where $\Phi(\mathbf{s}) = I\left(\|(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X} \mathbf{y}(\mathbf{s}) - \boldsymbol{\beta}^0(\mathbf{s})\|_2 > \frac{\varepsilon \sqrt{p}}{2}\right)$. Then by Lemma 5 (replacing ε by $\varepsilon/2$) we have

$$E_{P_0}[\Phi_{\boldsymbol{\beta}^1, \boldsymbol{\beta}^0}] \leq \exp\{-C_0 n_0\}, \quad E_{P_1}[\Phi_{\boldsymbol{\beta}^1, \boldsymbol{\beta}^0}] \leq \exp\{-C_0 n_0\},$$

where $C_0 > 0$ is a constant. Since $n_0 \geq c'n - 1$ for a positive constant c' , we have that $E_{P_0}[\Phi_{\boldsymbol{\beta}^1, \boldsymbol{\beta}^0}] \leq \exp\{-Cn\}$ and $E_{P_1}[\Phi_{\boldsymbol{\beta}^1, \boldsymbol{\beta}^0}] \leq \exp\{-Cn\}$. \square

Lemma 8. *There exists a test Ψ for testing $H_0 : \boldsymbol{\beta}(\mathbf{s}) = \boldsymbol{\beta}^0(\mathbf{s})$ against $H_1 : \boldsymbol{\beta}(\mathbf{s}) \in U_{\varepsilon, n}^c = U_\varepsilon^c \cap \mathbf{P} = \{\boldsymbol{\beta}(\mathbf{s}) \in \mathbf{P}_n : \|\boldsymbol{\beta}(\mathbf{s}) - \boldsymbol{\beta}^0(\mathbf{s})\| \geq \varepsilon\}$ in our proposed SVCM, such that*

$$E_{P_0}[\Psi] \leq \exp\{-d_0 n\}, \quad E_{P_1}[1 - \Psi] \leq \exp\{-d_1 n\},$$

for some constant d_0, d_1 with P_0 and P_1 corresponding to the probability distributions under H_0 and H_1 .

Proof. Let $\mathcal{N} = N(\frac{\varepsilon}{2}, \mathbf{P}_n, \|\cdot\|_\infty)$ be the covering number of \mathbf{P}_n by $\varepsilon/2$ -balls under the supreme norm. Then for all $\boldsymbol{\beta}(\mathbf{s}) \in U_{\varepsilon, n}^c$, we can find $\boldsymbol{\beta}^j(\mathbf{s}), j \in \{1, \dots, \mathcal{N}\}$ such that $\|\boldsymbol{\beta}^j(\mathbf{s}) - \boldsymbol{\beta}(\mathbf{s})\|_\infty \leq \frac{\varepsilon}{2}$, which implies that $\|\boldsymbol{\beta}^j(\mathbf{s}) - \boldsymbol{\beta}^0(\mathbf{s})\|_\infty \geq \|\boldsymbol{\beta}^0(\mathbf{s}) - \boldsymbol{\beta}(\mathbf{s})\|_\infty - \|\boldsymbol{\beta}^j(\mathbf{s}) - \boldsymbol{\beta}(\mathbf{s})\|_\infty \geq \frac{\varepsilon}{2}$ for all $j = 1, \dots, \mathcal{N}$. Following the notations and results in Lemma 7 with regard to $\varepsilon/2$, we have that the tests $\Phi_{\boldsymbol{\beta}^j, \boldsymbol{\beta}^0}$ all satisfy that $E_{\boldsymbol{\beta}^0}[\Phi_{\boldsymbol{\beta}^j, \boldsymbol{\beta}^0}] \leq \exp\{-d_1 n\}$ and $E_{\boldsymbol{\beta}^j}[\Phi_{\boldsymbol{\beta}^j, \boldsymbol{\beta}^0}] \leq \exp\{-d_1 n\}$ for some constant d_1 . Now for the test function

$$\Psi = \max_{j=1, \dots, \mathcal{N}} \Phi_{\boldsymbol{\beta}^j, \boldsymbol{\beta}^0},$$

which only depend on the set \mathbf{P}_n instead of specific $\boldsymbol{\beta}(\mathbf{s})$ in the alternative hypothesis,

$$E_{P_0}[\Psi] \leq \sum_{j=1}^{\mathcal{N}} E_{\boldsymbol{\beta}^j}[\Phi_{\boldsymbol{\beta}^j, \boldsymbol{\beta}^0}] \leq \mathcal{N} \exp\{-d_1 n\} < \exp\{Cpn^{\frac{d}{2\rho}} \varepsilon^{-d} - d_1 n\} \leq \exp\{-d_0 n\},$$

for some constant d_0 due to (28) and the fact that $n^{\frac{d}{2\rho}} = o(n)$. At the same time

$$E_{P_1}[1 - \Psi] \leq E_{\boldsymbol{\beta}^1}[1 - \Phi_{\boldsymbol{\beta}^1, \boldsymbol{\beta}^0}] \leq \exp\{-d_1 n\},$$

which complete our proof. \square

Now based on Lemma 1, equation (29) and Lemma 8, Theorem 2 follows from a direct application of Theorem A.1. of Choudhuri et al. (2004).

C Details about the MCMC algorithm

We list the details about our MCMC algorithm here. Denote by $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ the density function of $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We normally fix $v = w = 0.001$ in the inverse-gamma priors and fix $\theta^2 = 1$ for the local GPs.

- Updating $\tilde{\beta}(\mathbf{s}_i)$, $i = 1, \dots, n$: given the block structures of $\tilde{\beta}(\mathbf{s}_i)$, we update $\tilde{\beta}_k^g = \{\tilde{\beta}_k(\mathbf{s}_i)\}_{\mathbf{s}_i \in \mathcal{R}_g}$, $k = 1, \dots, p$, $g = 1, \dots, G$ separately with $p \times G$ M-H steps. Specifically, the full conditional $[\tilde{\beta}_k^g | \tilde{\beta}, \mathbf{u}, \sigma^2, \boldsymbol{\lambda}, \mathcal{Y}]$ is proportional to

$$h(\tilde{\beta}_k^g) = \left[\prod_{i: \mathbf{s}_i \in \mathcal{R}_g} \prod_{j=1}^m \phi \left(y_{j,-k}(\mathbf{s}_i); x_{jk} \tilde{\beta}_k(\mathbf{s}_i) I_{\lambda_k}[\tilde{\beta}_k(\mathbf{s}_i)], \sigma^2 \right) \right] \phi \left(\tilde{\beta}_k^g; \boldsymbol{\varphi}_g \mathbf{u}_k, \theta^2 K_g \right),$$

where $y_{j,-k}(\mathbf{s}_i) = y_j(\mathbf{s}_i) - \sum_{t \neq k} x_{jt} \tilde{\beta}_t(\mathbf{s}_i) I_{\lambda_k}[\tilde{\beta}_k(\mathbf{s}_i)]$. We adopt a Metropolis-Hasting (M-H) algorithm to update $\tilde{\beta}_k^g$ by first generating a proposal, $\tilde{\beta}_k^g + \Delta \tilde{\beta}_k^g$ with a zero mean Gaussian fluctuation $\Delta \tilde{\beta}_k^g$. Then we set $\tilde{\beta}_k^g \leftarrow \tilde{\beta}_k^g + \Delta \tilde{\beta}_k^g$ with probability: $\min \left\{ 1, \frac{h(\tilde{\beta}_k^g + \Delta \tilde{\beta}_k^g)}{h(\tilde{\beta}_k^g)} \right\}$.

- Updating σ^2 : draw σ^2 from its full conditional $[\sigma^2 | \tilde{\beta}, \boldsymbol{\lambda}, \mathcal{Y}]$ which is Inv-Ga($a_{\sigma^2}, b_{\sigma^2}$) where $a_{\sigma^2} = 0.001 + \frac{mn}{2}$ and $b_{\sigma^2} = 0.001 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m \left(y_j(\mathbf{s}) - \mathbf{x}_j^\top g_\lambda[\tilde{\beta}(\mathbf{s}_i)] \right)^2$.
- Updating $\boldsymbol{\lambda}$: we sequentially update $\lambda_1, \dots, \lambda_p$ with M-H algorithms. Specifically, for λ_k , the full conditional $[\lambda_k | \lambda_{-k}, \tilde{\beta}, \sigma^2, \mathcal{Y}]$ is

$$\hbar(\lambda_k) = \left[\prod_{i=1}^n \prod_{j=1}^m \phi \left(y_{j,-k}(\mathbf{s}), x_{jk} \tilde{\beta}_k(\mathbf{s}_i) I_{\lambda_k}[\tilde{\beta}_k(\mathbf{s}_i)], \sigma^2 \right) \right] \Pi(\lambda_k),$$

where $\Pi(\lambda_k)$ is the uniform empirical Bayes prior for λ_k defined in the previous section. The proposal for λ_k is generate from zero mean Gaussian fluctuations as $\lambda_k + \Delta \lambda_k$, which will be accepted with probability: $\min \left\{ 1, \frac{\hbar(\lambda_k + \Delta \lambda_k)}{\hbar(\lambda_k)} \right\}$.

- Updating $\{\mathbf{u}_k\}_{k=1}^p$: we sequentially update $\mathbf{u}_1, \dots, \mathbf{u}_p$ by drawing from their full conditionals $[\mathbf{u}_k | \tilde{\beta}, \tau_k^2]$. Specifically, we update \mathbf{u}_k by drawing from $N(\boldsymbol{\mu}_{\mathbf{u}_k}, \boldsymbol{\Sigma}_{\mathbf{u}_k})$ where $\boldsymbol{\mu}_{\mathbf{u}_k} = \boldsymbol{\Sigma}_{\mathbf{u}_k} \left(\theta^{-2} \sum_{g=1}^G \boldsymbol{\varphi}_g^\top K_g^{-1} \tilde{\beta}_k^g \right)$ and $\boldsymbol{\Sigma}_{\mathbf{u}_k} = \left(\sum_{g=1}^G \theta^{-2} \boldsymbol{\varphi}_g^\top K_g^{-1} \boldsymbol{\varphi}_g + \tau_k^{-2} \mathbf{Z}^{-1} \right)^{-1}$, with $\mathbf{Z} = \text{diag}(\zeta_1, \dots, \zeta_L)$.
- Updating $\{\tau_k^2\}_{k=1}^p$: we sequentially update $\tau_1^2, \dots, \tau_p^2$ by drawing from their full conditionals $[\tau_k^2 | \mathbf{u}_k]$. Specifically, we update τ_k^2 by drawing from Inv-Ga($a_{\tau_k^2}, b_{\tau_k^2}$) where $a_{\tau_k^2} = 0.001 + \frac{L}{2}$ and $b_{\tau_k^2} = 0.001 + \frac{1}{2} \mathbf{u}_k^\top \mathbf{Z}^{-1} \mathbf{u}_k$.
- Updating the spatial range parameter b within the SE kernel: this parameter can be updated by discretization. Specifically, within a reasonable range of b , we can calculate and store the dictionaries of $\boldsymbol{\varphi}_l$ and ζ_l , the kernel expansion results, with regard to each

discrete values of b on a grid basis. Then we can update b based on grid search within each MCMC iteration.

References

- Adler, R. J. and Taylor, J. E. (2009), *Random Fields and Geometry*, vol. 115, Springer.
- Armagan, A., Dunson, D. B., and Lee, J. (2013), “Generalized double Pareto shrinkage,” *Statistica Sinica*, 23, 119.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006), “Adaptive linear step-up procedures that control the false discovery rate,” *Biometrika*, 93, 491–507.
- Benjamini, Y. and Yekutieli, D. (2001), “The control of the false discovery rate in multiple testing under dependency,” *Annals of statistics*, 1165–1188.
- Bondell, H. D. and Reich, B. J. (2012), “Consistent high-dimensional Bayesian variable selection via penalized credible regions,” *Journal of the American Statistical Association*, 107, 1610–1624.
- Candes, E. and Tao, T. (2007), “The Dantzig selector: Statistical estimation when p is much larger than n ,” *The Annals of Statistics*, 2313–2351.
- Chan, T. F. and Shen, J. J. (2005), *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*, SIAM.
- Choudhuri, N., Ghosal, S., and Roy, A. (2004), “Bayesian estimation of the spectral density of a time series,” *Journal of the American Statistical Association*, 99, 1050–1059.
- Chumbley, J., Worsley, K. J., Flandin, G., and Friston, K. J. (2009), “False discovery rate revisited: FDR and topological inference using Gaussian random fields,” *NeuroImage*, 44, 62–70.
- Cressie, N. A. and Cassie, N. A. (1993), *Statistics for spatial data*, vol. 900, Wiley New York.
- Diggle, P. J., Tawn, J., and Moyeed, R. (1998), “Model-based geostatistics,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47, 299–350.
- Dunson, D. B. and Fox, E. B. (2012), “Multiresolution gaussian processes,” in *Advances in Neural Information Processing Systems*, pp. 737–745.
- Fan, J. and Li, R. (2001), “Variable selection via nonconcave penalized likelihood and its oracle properties,” *Journal of the American Statistical Association*, 96, 1348–1360.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. (1995), “Statistical parametric maps in functional imaging: a general linear approach,” *Human brain mapping*, 2, 189–210.
- Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003), “Spatial modeling with spatially varying coefficient processes,” *Journal of the American Statistical Association*, 98, 387–396.

- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002), “Thresholding of statistical maps in functional neuroimaging using the false discovery rate,” *Neuroimage*, 15, 870–878.
- George, E. I. and McCulloch, R. E. (1993), “Variable selection via Gibbs sampling,” *Journal of the American Statistical Association*, 88, 881–889.
- Ghosal, S., Roy, A., et al. (2006), “Posterior consistency of Gaussian process prior for non-parametric binary regression,” *The Annals of Statistics*, 34, 2413–2429.
- Hadjikhani, N., Joseph, R. M., Snyder, J., Chabris, C. F., Clark, J., Steele, S., McGrath, L., Vangel, M., Aharon, I., Feczko, E., et al. (2004), “Activation of the fusiform gyrus when individuals with autism spectrum disorder view faces,” *Neuroimage*, 22, 1141–1150.
- Hans, C. (2009), “Bayesian lasso regression,” *Biometrika*, 96, 835–845.
- Hoeffding, W. (1963), “Probability inequalities for sums of bounded random variables,” *Journal of the American Statistical Association*, 58, 13–30.
- Ishwaran, H. and Rao, J. S. (2005), “Spike and slab variable selection: frequentist and Bayesian strategies,” *Annals of Statistics*, 730–773.
- Johnson, V. E. and Rossell, D. (2012), “Bayesian model selection in high-dimensional settings,” *Journal of the American Statistical Association*, 107, 649–660.
- Lazar, N. (2008), *The statistical analysis of functional MRI data*, Springer.
- Li, F. and Zhang, N. R. (2010), “Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics,” *Journal of the American Statistical Association*, 105.
- Li, Y., Zhu, H., Shen, D., Lin, W., Gilmore, J. H., and Ibrahim, J. G. (2011), “Multiscale adaptive regression models for neuroimaging data,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73, 559–578.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008), “Mixtures of g priors for Bayesian variable selection,” *Journal of the American Statistical Association*, 103.
- Mitchell, T. J. and Beauchamp, J. J. (1988), “Bayesian variable selection in linear regression,” *Journal of the American Statistical Association*, 83, 1023–1032.
- Narisetty, N. N., He, X., et al. (2014), “Bayesian variable selection with shrinking and diffusing priors,” *The Annals of Statistics*, 42, 789–817.
- Nichols, T. and Hayasaka, S. (2003), “Controlling the familywise error rate in functional neuroimaging: a comparative review,” *Statistical methods in medical research*, 12, 419–446.
- Park, T. and Casella, G. (2008), “The bayesian lasso,” *Journal of the American Statistical Association*, 103, 681–686.

- Polson, N. G., Scott, J. G., and Windle, J. (2014), “The Bayesian Bridge,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.
- Polzehl, J. and Spokoiny, V. G. (2000), “Adaptive weights smoothing with applications to image restoration,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62, 335–354.
- Qiu, P. (2007), “Jump surface estimation, edge detection, and image restoration,” *Journal of the American Statistical Association*, 102, 745–756.
- Reich, B. J., Fuentes, M., Herring, A. H., and Evenson, K. R. (2010), “Bayesian variable selection for multivariate spatially varying coefficient regression,” *Biometrics*, 66, 772–782.
- Smith, D. L., Lucey, B., Waller, L. A., Childs, J. E., and Real, L. A. (2002), “Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes,” *Proceedings of the National Academy of Sciences*, 99, 3668–3672.
- Smith, M. and Fahrmeir, L. (2007), “Spatial Bayesian variable selection with application to functional magnetic resonance imaging,” *Journal of the American Statistical Association*, 102, 417–431.
- Smith, M., Pütz, B., Auer, D., and Fahrmeir, L. (2003), “Assessing brain activity through spatial Bayesian variable selection,” *NeuroImage*, 20, 802–815.
- Stein, M. L. (1999), *Interpolation of spatial data: some theory for kriging*, Springer.
- Tabelow, K., Polzehl, J., Spokoiny, V., and Voss, H. U. (2008a), “Diffusion tensor imaging: structural adaptive smoothing,” *NeuroImage*, 39, 1763–1773.
- Tabelow, K., Polzehl, J., Ulug, A. M., Dyke, J. P., Watts, R., Heier, L. A., and Voss, H. U. (2008b), “Accurate localization of brain activity in presurgical fmri by structure adaptive smoothing,” *Medical Imaging, IEEE Transactions on*, 27, 531–537.
- Tibshirani, R. (1996), “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002), “Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain,” *Neuroimage*, 15, 273–289.
- Van Der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer.
- Wang, J., Zhu, H., Fan, J., Giovanello, K., and Lin, W. (2013), “Multiscale adaptive smoothing models for the hemodynamic response function in fMRI,” *The annals of applied statistics*, 7, 904.
- Williams, C. K. and Rasmussen, C. E. (2006), “Gaussian processes for machine learning,” *the MIT Press*, 2, 4.

- Worsley, K. J., Taylor, J. E., Tomaiuolo, F., and Lerch, J. (2004), “Unified univariate and multivariate random field theory,” *NeuroImage*, 23, 189–195.
- Xie, Y. and Carlin, B. P. (2004), “A Note on Bayesian Learning and Identifiability in Hierarchical Models,” .
- Yue, Y., Loh, J. M., and Lindquist, M. A. (2010), “Adaptive spatial smoothing of fMRI images,” *Statistics and its Interface*, 3, 3–13.
- Zhu, H., Fan, J., and Kong, L. (2014), “Spatially Varying Coefficient Model for Neuroimaging Data with Jump Discontinuities,” *Journal of the American Statistical Association*.
- Zou, H. (2006), “The adaptive lasso and its oracle properties,” *Journal of the American statistical association*, 101, 1418–1429.
- Zou, Q.-H., Zhu, C.-Z., Yang, Y., Zuo, X.-N., Long, X.-Y., Cao, Q.-J., Wang, Y.-F., and Zang, Y.-F. (2008), “An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: fractional ALFF,” *Journal of neuroscience methods*, 172, 137–141.